

# EVOLUTION AND DEVELOPMENT OF STORAGE ROOTS IN MORNING GLORIES

(CONVOLVULACEAE)

by

LAUREN ASHLEY ESERMAN

(Under the Direction of Jim Leebens-Mack)

## ABSTRACT

Storage roots are an important adaptation to harsh environmental conditions. The number of plant species with storage roots is not known, likely because storage roots are difficult to study. Many species distributed across the morning glory family, Convolvulaceae, form storage roots, including sweetpotato, *Ipomoea batatas* (L.) Lam. I employed a comparative approach to investigate the evolution and development of storage roots in morning glories. We first estimated relationships among major morning glory lineages using plastome sequences to examine the evolution of three ecologically important traits: storage roots, flower color and ergot alkaloid presence. We then used target enrichment to estimate relationships in the sweetpotato complex, as well as the timing and extent of hybridization. While accounting for phylogenetic relatedness among species, we tested for a correlation between polyploidy and root traits in the Batatas complex. Finally, we examined anatomical and transcriptomic changes associated with storage root formation in two pairs of distantly related morning glory species. These findings suggest numerous independent origins of storage roots throughout morning glory evolution. Within the Batatas complex, phylogenomic analyses revealed ancient hybridization with minimal evidence for ongoing gene flow. In addition to the possibility that hybridization among unrelated lineages

has led to introgression of loci controlling storage root formation and the origin of storage roots, polyploidy may have also played a role. This hypothesis was tested, and we found that ploidy level and genome size were poor predictors of storage root formation. Therefore, factors other than whole genome duplications are needed to explain root trait diversity in the Batatas complex. Finally, comparative anatomical and transcriptomic analyses revealed that storage roots of sweetpotato and *Merremia dissecta*, two distantly related morning glory species, utilize a common core set of genes in storage root formation despite exhibiting different storage root developmental patterns. Many of the genes showing increased expression during storage root formation are involved in the starch biosynthesis and others are regulators of starch synthesis and cambium formation. Taken together, the results support a multifaceted picture of storage root evolution and development, suggesting this is a complex morphological trait with numerous evolutionary origins.

INDEX WORDS: Morning glories, *Ipomoea*, storage roots, sweetpotato, crop wild relatives, hybridization, polyploidy, phylogenomics

EVOLUTION AND DEVELOPMENT OF STORAGE ROOTS IN MORNING GLORIES  
(CONVOLVULACEAE)

by

LAUREN ASHLEY ESERMAN

BS, Southeastern Louisiana University, 2009

MS, Southeastern Louisiana University, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial  
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

© 2017

Lauren Ashley Eserman

All Rights Reserved

EVOLUTION AND DEVELOPMENT OF STORAGE ROOTS IN MORNING GLORIES

(CONVOLVULACEAE)

by

LAUREN ASHLEY ESERMAN

Major Professor:	Jim Leebens-Mack
Committee:	Shu-Mei Chang
	Russell Malmberg
	Chung-Jui Tsai
	G. Craig Yencho

Electronic Version Approved:

Suzanne Barbour  
Dean of the Graduate School  
The University of Georgia  
August 2017

## ACKNOWLEDGEMENTS

I would first like to thank my advisor, Jim Leebens-Mack, who has always encouraged me to think creatively and deeply about science. I would also like to thank my committee, G. Craig Yencho, Shumei Chang, CJ Tsai, and Russell Malmberg, for their always insightful conversations. Bob Jarret at USDA provided intellectual support and funding for parts of this research. I also owe a great deal of gratitude to Rick Miller, who introduced me to morning glories and, thus, a lasting fascination with these incredible plants.

Kevin Turner and Mike Boyd always took great care of my plants. Magdy Alabady provided assistance with RNA isolations and library preparation. Beth Richardson, Brigitte Bruns, and Zheng-Hua Ye helped with microscopy. Charmi Patel measured root traits. Raj Ayyampalayam and Karolina Heyduk assisted with bioinformatics. The Batatas Complex research was supported by the USDA and a National Science Foundation Doctoral Dissertation Improvement Grant (NSF DEB 1601251). Funding for transcriptome sequencing came from a Palfrey award from the Plant Biology Department at UGA, a Rosemary Grant Award from the Society for the Study of Evolution, a Grant in Aid of Research from the Society for Integrative and Comparative Biology and the USDA.

Finally, I am eternally grateful for my parents, grandparents, family, friends, and my ever encouraging boyfriend, Brandon Campbell. None of this would be possible without their unwavering love and support.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW .....	1
The Convolvulaceae family .....	1
The Batatas Complex .....	2
Factors influencing the evolution of underground storage organs.....	5
Developmental biology and genetic control of storage root formation .....	7
Agricultural importance of storage roots .....	8
Confusion in terminology .....	8
References.....	9
2 PHYLOGENETICS AND DIVERSIFICATION OF MORNING GLORIES (TRIBE IPOMOEAE, CONVULVULACEAE) BASED ON WHOLE PLASTOME SEQUENCES. ....	14
Abstract .....	15
Introduction.....	16
Materials and Methods.....	20
Results.....	26

Discussion.....	29
References.....	40
3 THE EVOLUTIONARY HISTORY OF THE SWEETPOTATO COMPLEX, <i>IPOMOEA</i> SERIES <i>BATATAS</i> (CHOISY) D. F. AUSTIN, USING TARGET ENRICHMENT .....	53
Abstract.....	54
Introduction.....	54
Methods.....	58
Results.....	63
Discussion.....	67
References.....	72
4 EVOLUTION OF POLYPLOIDY AND STORAGE ROOTS IN SWEETPOTATO AND ITS WILD RELATIVES ( <i>IPOMOEA</i> SERIES <i>BATATAS</i> (CHOISY) D. F. AUSTIN).....	89
Abstract.....	90
Introduction.....	90
Methods.....	93
Results.....	101
Discussion.....	107
References.....	113
5 PARALLEL EVOLUTION OF STORAGE ROOTS IN MORNING GLORIES (CONVOLVULACEAE).....	141
Abstract.....	142

Introduction.....	143
Methods.....	145
Results.....	150
Discussion.....	153
References.....	157
6 CONCLUSION AND DISCUSSION .....	170
Morning glory phylogenomics.....	170
Storage root evolution.....	171
References.....	173
APPENDICES	
A SUPPLEMENTARY FIGURES AND TABLES – CHAPTER II .....	175
B SUPPLEMENTARY FIGURES AND TABLES – CHAPTER IV .....	186

## LIST OF TABLES

	Page
Table 2.1: Mean, minimum and maximum node ages of clades denoted in figure 2.2. ....	49
Table 3.1: Accession information for individuals included in analyses in this study.....	78
Table 3.2: Summary of genome size measurements, standard deviation from three flow cytometric measurements for each sample, and inferred ploidy for the accessions measured in this study.....	79
Table 3.3: Summary of sequencing information for the accessions sequenced in this study. ....	80
Table 3.4: Alignment statistics for the five datasets used in analyses in this study. ....	81
Table 4.1: Accession information for the individuals sequenced in this study. ....	119
Table 4.2: Genome size measurements for accessions used in this study. ....	121
Table 4.3: Between experiment controls for flow cytometric measurement of genome size. Shown are genome size measurements and inferred ploidy level. ....	123
Table 4.4: Across year controls for flow cytometric measurement of genome size.....	124
Table 4.5: Sequencing results for all accessions sequenced in this project. ....	125
Table 4.6: Alignment statistics for the two datasets used in this study. ....	127
Table 4.7: Proportion of variance in root traits explained by the five phylogenetic principal component axes for the pPCA that did not include starch measurements.....	128
Table 4.8: Proportion of variance in root traits explained by the five phylogenetic principal component axes for the pPCA that included starch measurements. ....	129
Table 4.9: Measures of phylogenetic signal for the traits used in this study. ....	130

Table 4.10: Results of the linear regression of phylogenetic independent contrasts of log-transformed genome size and the first two axes of the phylogenetic principal components analysis.....	131
Table 5.1: Transcriptome assembly statistics. ....	162
Table 5.2: Assembly statistics after successive filtering by IsoPct and FPKM, Swiss-prot annotations, and Decon-Seq.....	163
Table 5.3: Top ten most abundant gene ontology (GO) categories represented in genes differentially expressed between storage roots and fine roots of <i>Ipomoea batatas</i> and <i>Merremia dissecta</i> considering each species separately.....	164
Table 5.4: Top ten most abundant gene ontology (GO) categories represented in the set of orthologous genes differentially expressed between storage roots and fine roots in both <i>Ipomoea batatas</i> and <i>Merremia dissecta</i> . ....	165
Table S2.1: Voucher, GenBank accession, and locality information for species included in this study. ....	175
Table S2.2: Species used in this analysis, accession number, sequencing technology used for a particular accession, number of raw reads, and chloroplast genome coverage. ....	177
Table S2.3: Species used in this analysis with GC content and placement of inverted repeat boundaries. ....	178
Table S2.4: Table of characters, character states, and references of each character state used in the likelihood-based ancestral character state reconstructions. ....	179

## LIST OF FIGURES

	Page
Figure 2.1: Chloroplast genome of <i>Ipomoea hederifolia</i> .....	50
Figure 2.2: Phylogeny of the Ipomoeae based on whole chloroplast genome sequences.....	51
Figure 2.3: Results of the divergence time analysis. ....	52
Figure 3.1: Phylogenetic relationships among the diploid wild relatives of sweetpotato. ....	82
Figure 3.2: Phylogenetic placement of hexaploid cultivated sweetpotato, cv. Tinian (a,b) and tetraploid <i>Ipomoea tabascanana</i> (c,d). ....	83
Figure 3.3: Results of the HyDe analysis.....	84
Figure 3.4: Results from the PhyloNet analysis.....	85
Figure 3.5: Histograms depicting pairwise genetic distance (uncorrected p-distance) between taxa designated as hybrids and the inferred parental taxa from the HyDe analysis. ....	86
Figure 3.6: Results of the HyDe analysis which included hexaploid <i>Ipomoea batatas</i> (a,b) and <i>I.</i> <i>tabascanana</i> (c,d).....	87
Figure 3.7: Results from the PhyloNet analyses which included <i>Ipomoea batatas</i> (a) and <i>I.</i> <i>tabascanana</i> (b). ....	88
Figure 4.1: Phylogeny of the Batatas complex estimated in ASTRAL-II including only diploid individuals.....	132
Figure 4.2: Phylogeny of the Batatas complex estimated in ASTRAL-II including diploid and polyploid individuals. ....	133
Figure 4.3: Evolutionary history of polyploidy in the Batatas complex.....	134

Figure 4.4: Phylogeny of the Batatas complex estimated in ASTRAL-II illustrating results from the HyDe analysis of diploid taxa.....	135
Figure 4.5: Phylogeny of the Batatas complex estimated in ASTRAL-II illustrating results from the HyDe analysis of diploid and polyploid taxa.....	136
Figure 4.6: Results of the PhyloNet analysis of five random samples of diploid taxa. ....	137
Figure 4.7: Phylogeny of the Batatas complex with genome size and root trait values. ....	138
Figure 4.8: Biplot of phylogenetic principal components analyses for the dataset which did not include starch (a, c) and the dataset that did include starch measurements (b, d). ....	139
Figure 4.9: Linear regression between phylogenetic independent contrasts of the first PC axis (PC1) and log-transformed genome size in datasets which did not include starch (a) and did include starch (b) in the phylogenetic principal components analysis. ....	140
Figure 5.1: Root cross sections from three pairs of species, where one member of the species pair forms storage roots and the other does not. ....	166
Figure 5.2: Heat map of genes differentially expressed between storage roots and fine roots of sweetpotato, <i>Ipomoea batatas</i> (a) and <i>Merremia dissecta</i> (b). ....	167
Figure 5.3: Starch biosynthetic pathway adapted from Bahaji et al. 2014. ....	168
Figure 5.4: Mean TMM-normalized FPKM values for the seven transcription factors found to be significantly differentially expressed between storage and fine roots in both <i>Ipomoea batatas</i> and <i>Merremia dissecta</i> at a FDR <0.05. ....	169
Figure S2.1: Ancestral character state reconstructions of (a) root architecture, (b) flower color, and (c) ergot alkaloid presence and table of character states used in the analyses. ....	184
Figure S2.2: Phylogeny of the Ipomoeae based on 82 protein coding and rRNA sequences....	185

Figure S4.1: Phylogeny of the Batatas complex estimated in SVDQuartets including only diploid individuals.....187

Figure S4.2: Phylogeny of the Batatas complex estimated using a RAxML analysis of the concatenated gene matrix, which included only diploid.....188

Figure S4.3: Phylogeny of the Batatas complex estimated in SVDQuartets including diploid and polyploid individuals. ....189

Figure S4.4: Phylogeny of the Batatas complex estimated using a RAxML analysis of the concatenated gene matrix, which included diploid and polyploid individuals. ....190

## CHAPTER I:

### INTRODUCTION AND LITERATURE REVIEW

#### **The Convolvulaceae family**

The plant family Convolvulaceae is a large, diverse group of angiosperms consisting of approximately 1600-1700 species across 55-60 genera (Mabberley 2008). Species in the Convolvulaceae are primarily twining vines but holoparasites (*Cuscuta*), lianas (*Ipomoea ampullacea*, *Merremia discoidesperma*), shrubs (*I. stans*, *I. leptophylla*), and trees (*I. series Arborescentes*) are represented (Mabberley 2008). The Convolvulaceae includes both annual and perennial species (McDonald 1994; Austin 1998). Members of this family are found throughout the world but are primarily concentrated in tropical and subtropical regions (McDonald 1991; Austin and Huáman 1996; Austin 1998).

There is a great degree of morphological diversity in the Convolvulaceae, and a suite of ecologically important phenotypes vary greatly across the morning glory phylogeny (Manos et al. 2001). Morning glories have primarily been studied as a model system for understanding the evolution of plant mating systems and the genetic basis and evolution of flower color, specifically the anthocyanin biosynthetic pathway (reviewed in Baucom et al. 2011). Furthermore, morning glories have been critical in understanding plant-endosymbiont relationships, as morning glories are one of two plant groups to form an association with ergot alkaloid producing fungi (Kucht et al. 2004; Steiner et al. 2011).

One ecologically important trait which has received little attention to date in morning glories is the storage root. There is no current estimate of the number of morning glory species which produce storage roots, likely because storage roots are challenging to locate and identify. Often, roots are not examined when field biologists collect wild plants. This is because either permits were not issued to collect below-ground tissue or the storage root was difficult to find, as some storage roots are large and form deep in the soil, e.g. *Ipomoea pandurata* storage roots can be 2 m in length, weigh up to 30 kg, and are buried deep below the soil surface (Horak and Wax 1991). Thus, most herbarium specimens do not include storage root tissue.

A more careful evaluation of the diversity in storage root formation in morning glories is critical because many storage roots are of agronomic or medical importance. Storage roots of many morning glories were used for medicinal purposes by people native to the Neotropics, primarily as purgatives (McDonald 1989; Linajes et al. 1994). Furthermore, sweetpotato, which was first domesticated in the Americas (Roullier et al. 2011), is agriculturally important worldwide for its carbohydrate and vitamin-A rich storage roots (Hotz et al. 2012). A thorough understanding of storage root formation and variation across morning glories can aid in sweetpotato breeding. Therefore, the major focus of the second chapter of my dissertation is reconstructing the evolutionary relationships in the Ipomoeae, estimating divergence times among lineages, and characterizing the evolution of ecologically important plant traits, such as storage root formation, flower color, and ergot alkaloid presence.

### **The Batatas Complex**

The Batatas Complex, *Ipomoea* series *Batatas* (Choisy) D. F. Austin, contains the cultivated storage root crop sweetpotato and its wild relatives. This group of species shows variation in storage root formation, where sweetpotato forms large storage roots and some but

not all wild relatives are reported to form smaller storage roots (Austin 1978; McDonald 1994; Komaki and Katayama 1999), making this an ideal group in which to study the evolution of storage roots. However, there are several factors complicating an examination of the evolution of storage roots in the Batatas complex, namely, polyploidy and hybridization.

Polyploidy is an increase in the number of sets of chromosomes in an organism (Ramsey and Schemske 1998; Otto and Whitton 2000). Polyploid organisms are formed either through auto- or allopolyploidization or some combination of the two. Autopolyploids are the result of chromosome doubling and are often the result of a cross between unreduced gametes within an individual (Ramsey and Schemske 1998). Allopolyploids, however, are hybrids between either different populations of the same species or between two separate species without a reduction in chromosome number (Ramsey and Schemske 1998). Several studies of the Batatas complex have reported a high frequency of polyploid individuals across multiple different species (Ozias-Akins and Jarret 1994; Roullier et al. 2013b). If polyploids are found to exist at a high frequency in the Batatas complex, this may complicate inferences of trait evolution, as polyploids can be problematic for phylogenetic reconstruction. After polyploidization, more than two chromosome copies exist, allowing for the potential for divergence among gene copies. Therefore, when we sequence loci in polyploid taxa, we may expect to see multiple copies of a particular gene. Following polyploidization, genome fractionation can occur, where a subset of genes in the genome are retained in single copy (Feldman et al. 1997; Wendel 2015). Fractionation can also occur in a biased fashion, where a particular subset of genes are retained in single copy more so than others (Schnable et al. 2011; Cheng et al. 2012). Phylogenomic analysis and subsequent inferences of trait evolution would be more straightforward if we find that genomic fractionation occurs in a more biased manner in Batatas complex polyploids because we could consistently

have a set of genes retained in single copy for phylogenomics. However, determining whether fractionation has occurred is difficult, especially if the parents of the polyploid are unknown.

Hybridization, whether homoploid or polyploid, also has implications for our inferences of trait evolution. In homoploid hybridization, individuals from two separately evolving lineages produce offspring whose chromosomes contain portions from both parents (Abbott et al. 2016). In contrast, polyploid hybrids or allopolyploids, contain full chromosome sets from both parents (Otto and Whitton 2000). Often, allopolyploids are reproductively isolated from the parental species (Mallet 2007). In contrast, diploid hybrids can vary from fully interfertile with parental taxa, viable but reproductively isolated from parental species, sterile, and even inviable (Mallet 2007). The genomics and phenotypic effects of recent homoploid hybridization have been studied in great detail and have revealed that hybrid phenotypes vary from parental traits in a number of ways. For polygenic traits, hybrids sometimes exhibit morphologies intermediate between both parental taxa (Grant and Grant 1994). In some plant lineages, hybrids exhibit heterosis or hybrid vigor, where hybrids exhibit greater overall growth than either parent (Birchler et al. 2003). In other cases, hybridization among more distantly related lineages can result in separate species appearing morphologically similar due to introgression of loci controlling the phenotypic traits of interest (The Heliconius Genome Consortium et al. 2012).

In all of these cases, it is clear that the phenotypic consequences of hybridization are highly dependent upon the genetics of the parents in the hybridization event. Yet most work on phenotypic evolution in hybrid lineages has focused on recent hybrid species, especially contemporary hybrid zones. In contrast, the phenotypic effects of ancient hybridization followed by speciation have received relatively less attention likely because this question is much more difficult to answer. While the fate of genes in a more recent hybridization are more clearly linked

to parental phenotypic evolution, an ancient hybridization event is followed by diversification over long time periods. The fate of genes in ancient events is compounded by the time these genomic regions have to independently evolve from parental copies. After the initial hybridization event, different portions of the genome may undergo varying degrees of selection and result in different signals of parental ancestry. Studies of genomic stabilization following hybridization have found that regions of genome subject to more purifying selection are typically fixed for one or the other parental alleles, but regions of the genome under more relaxed selective constraint tend to show a signature of hybrid ancestry (Sankaraman et al. 2015; Schumer et al. 2016). At a broad level, these results suggest that some but not all of the genome following hybridization will retain the signature of hybrid ancestry. Furthermore, if we observe ancient hybridization followed by speciation, there are no available evolutionary models which have been developed to model the evolution of traits following hybridization.

In the third chapter of this dissertation, I explore the evolutionary history of the Batatas complex. This chapter examines both the influence of incomplete lineage sorting and hybridization on gene tree discordance in the Batatas complex. Furthermore, I characterize the timing of hybridization, testing whether hybridization is ongoing or ancient.

### **Factors influencing the evolution of underground storage organs**

Geophytes are plants that store carbohydrates long-term in large underground storage organs (Dafni et al. 1981). These underground storage organs can be comprised of root (e.g. storage roots), stem (e.g. corms, rhizomes, tubers), or leaf tissue (e.g. bulbs). Plants can store different types of carbohydrates in underground storage organs, and while the primary carbohydrate being stored can vary by species, plants tend to store either simple sugars, e.g. sugar beet (Giaquinta 1979), or starch, e.g. sweetpotato and cassava (Cervantes-Flores et al.

2010; Mejia-Aguero et al. 2012). Geophytic plants have been shown to continuously allocate carbohydrates to underground reserves throughout the growing season; whereas, non-geophytes tend to allocate carbohydrates to below-ground tissue only during the first few stages of growth after which allocation ceases (De Souza and Viera Da Silva 1987; Ruiters et al. 1993; Ruiters and McKenzie 1994).

Numerous hypotheses have been proposed to explain the diversity of plant species with underground storage organs, particularly storage roots. In his 1880 book, *The Power of Movement in Plants*, Darwin (1880) suggest that the geophytic habit is associated with a reduction in cotyledon size. This idea was echoed by Sargent (1904), who furthered this hypothesis by suggesting that both cotyledon size reduction and underground carbohydrate storage are adaptations to harsh environmental conditions. The idea was related to plant economics, where reduced allocation to cotyledon growth also meant increased allocation to growth and expansion of the radicle in a seedling (Sargent 1904). Others have suggested that increases in genome size, due to polyploidy or other causes such as retrotransposon proliferation, have driven the evolution of the geophytic habit (Grime and Mowforth 1985; Veselý et al. 2012, 2013). Polyploidy has also been tied to traits ecologically linked to a geophytic habit such as perenniality (Tank and Olmstead 2008). Underground carbohydrate storage is almost certainly an adaptation to harsh environmental conditions (Sargent 1904; Dafni et al. 1981; Bell et al. 1996). Many plants cope with environmental stresses by having underground storage organs. When environmental conditions are unfavorable, such vegetation die-off following fire, the ability to resprout from underground carbohydrate reserves has advantages (Bell et al. 1996). Additionally, some species allocate more starch to root reserves during drought conditions, in preparation for regrowth following the drought period (Galvez et al. 2011)

In the fourth chapter of my dissertation, I continue to explore the evolutionary history of the Batatas complex but using a larger sample of taxa. Further, I am testing for a possible correlation between genome size and root traits to test the hypothesis that whole genome duplications led to the large storage roots in sweetpotato.

### **Developmental biology and genetic control of storage root formation**

Despite the clear ecological and economic significance of storage roots, there is a paucity of information on the development and genetic control of storage root formation. Much of what is known about the developmental biology and genetic control of underground carbohydrate storage comes from studies of potato, sweetpotato, and sugar beet. In sweetpotato, a fine lateral root will transition into a storage root by accumulation of starch in tissue known as anomalous cambium (Artschwager 1924; Wilson and Lowe 1973; Lowe and Wilson 1974a, 1974b; Firon et al. 2013). Comparisons of transcriptomes between storage roots and fine roots in sweetpotato have demonstrated three main results: (1) genes involved in lignin biosynthesis are downregulated in storage roots, (2) genes involved in starch biosynthesis are upregulated in storage roots, and (3) many transcription factors are differentially regulated between storage roots and fine roots (You et al. 2003; Firon et al. 2013). In addition, three genes have been experimentally shown to regulate storage root formation in sweetpotato. One is an alpha-expansin gene, and downregulation of this gene in fine roots results in shorter cells, consistent with what is seen in starch accumulating cells in storage roots (Noh et al. 2013). The other two are MADS-box genes, and overexpression of these genes in both sweetpotato and potato fine roots results in root swelling (Noh et al. 2010, 2013). Taken together, these studies suggest that storage root formation is a complex process involving action from a number of biosynthetic pathways as well as careful control by transcription factors.

Therefore, in the fifth dissertation chapter, I use a comparative approach to characterize the anatomical and transcriptomic changes associated with storage root formation in pairs of species where one forms storage roots and the other does not. Species pairs were sampled from two separate tribes in the morning glory family Convolvulaceae.

### **Agricultural importance of storage roots**

Of the myriad of underground storage organs, storage roots are important both ecologically and economically. In total, root and tuber crops in total constituted over 800 million tonnes of crop produced. In this category, cassava and sweetpotato are the highest producing root crops. In 2014 alone, there were over 100 million tonnes of sweetpotato crop produced (FAO 2016). Because of its high beta carotene content, sweetpotato has also been critical in battling vitamin A deficiency in underdeveloped regions of the world (Hotz et al. 2012). Furthermore, this crop performs well even under adverse environmental conditions, such as limited water (van Heerden and Laurie 2008; Andrade et al. 2016).

### **Confusion in terminology**

In the literature, the terms tuber, root tuber and storage root are often applied to describe the same morphological feature. This likely arises from the fact that the term tuber has been used to describe thickened roots in botanical literature going back over two centuries (e.g. Lamarck 1791). Tuberization is often synonymous with storage root formation in the literature (Ku et al. 2008; Nedunchezhiyan et al. 2012). The term tuber is most often used to describe an underground storage stem, such as potato, rather than a root. In this dissertation, the term tuber is used throughout Chapter II because the confusion with this term was not realized until after publication. Therefore, in chapters III, IV, and V, I use the more specific term “storage root” rather than tuber.

## References

- Abbott R.J., Barton N.H., Good J.M. 2016. Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* 25:2325–2332.
- Andrade M.I., Naico A., Ricardo J., Eyzaguirre R., Makunde G.S., Ortiz R., Grüneberg W.J. 2016. Genotype × environment interaction and selection for drought adaptation in sweetpotato (*Ipomoea batatas* [L.] Lam.) in Mozambique. *Euphytica*. 209:261–280.
- Artschwager E. 1924. On the anatomy of the sweet potato with notes on internal breakdown. *J. Agric. Res.* 27:157–166.
- Austin D.F. 1978. The *Ipomoea batatas* complex. I. Taxonomy. *Bull. Torrey Bot. Club.* 105:114–129.
- Austin D.F. 1998. Convolvulaceae- Morning Glory Family. *J. Arizona-Nevada Acad. Sci.* 30:61–83.
- Austin D.F., Huáman Z. 1996. A Synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon.* 45:3–38.
- Baucom R.S., Chang S., Kniskern J.M., Rausher M.D., Stinchcombe J.R. 2011. Morning glory as a powerful model in ecological genomics: tracing adaptation through both natural and artificial selection. *Heredity.* 107:377–85.
- Bell T.L., Pate J.S., Dixon K.W. 1996. Relationships Between Fire Response, Morphology, Root Anatomy and Starch Distribution in South-west Australian Epacridaceae. *Ann. Bot.* 77:357–364.
- Birchler J.A., Auger D.L., Riddle N.C. 2003. In Search of the Molecular Basis of Heterosis. *Plant Cell.* 15:2236–2239.
- Cervantes-Flores J.C., Sosinski B., Pecota K. V., Mwangi R.O.M., Catignani G.L., Truong V.D., Watkins R.H., Ulmer M.R., Yencho G.C. 2010. Identification of quantitative trait loci for dry-matter, starch, and β-carotene content in sweetpotato. *Mol. Breed.* 28:201–216.
- Cheng F., Wu J., Fang L., Sun S., Liu B., Lin K., Bonnema G., Wang X. 2012. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One.* 7.
- Dafni A., Cohen D., Noy-Mier I. 1981. Life-Cycle Variation in Geophytes. *Ann. Missouri Bot. Gard.* 68:652–660.
- Darwin C. 1880. *The power of movement in plants*. London: John Murray.
- FAO. 2016. Food and Agriculture Organization of the United Nations. .
- Feldman M., Liu B., Segal G., Abbo S., Levy A.A., Vega J.M. 1997. Rapid elimination of low-

- copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics*. 147:1381–1387.
- Firon N., LaBonte D., Villordon A., Kfir Y., Solis J., Lapis E., Perlman T.S., Doron-Faigenboim A., Hetzroni A., Althan L., Adani Nadir L. 2013. Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics*. 14:460.
- Galvez D.A., Landhäusser S.M., Tyree M.T. 2011. Root carbon reserve dynamics in aspen seedlings: Does simulated drought induce reserve limitation? *Tree Physiol.* 31:250–257.
- Giaquinta R.T. 1979. Sucrose translocation and storage in the sugar beet. *Plant Physiol.* 63:828–832.
- Grant P.R., Grant B.R. 1994. Phenotypic and Genetic Effects of Hybridization in Darwin's Finches. *Evolution*. 48:297–316.
- Grime J.P., Mowforth M.A. 1985. Variation in genome size – an ecological interpretation. *Nature*. 299:151–153.
- van Heerden P.D.R., Laurie R. 2008. Effects of prolonged restriction in water supply on photosynthesis, shoot development and storage root yield in sweet potato. *Physiol. Plant.* 134:99–109.
- Horak M.J., Wax L.M. 1991. Growth and Development of Bigroot Morningglory (*Ipomoea pandurata*). *Weed Technol.* 5:805–810.
- Hotz C., Loechl C., de Brauw A., Eozenou P., Gilligan D., Moursi M., Munhaua B., van Jaarsveld P., Carriquiry A., Meenakshi J. V. 2012. A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *Br. J. Nutr.* 108:163–76.
- Komaki K., Katayama K. 1999. Root thickness of diploid *Ipomoea trifida* (H. B. K.) G. Don and performance of progeny derived from the cross with sweetpotato. *Breed. Sci.* 49:123–129.
- Ku A.T., Huang Y.-S., Wang Y.-S., Ma D., Yeh K.-W. 2008. IbMADS1 (*Ipomoea batatas* MADS-box 1 gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). *Ann. Bot.* 102:57–67.
- Kucht S., Gross J., Hussein Y., Grothe T., Keller U., Basar S., König W.A., Steiner U., Leistner E. 2004. Elimination of ergoline alkaloids following treatment of *Ipomoea asarifolia* (Convolvulaceae) with fungicides. *Planta*. 219:619–25.
- Lamarck J.-B. 1791. Tableau encyclopédique et méthodique des trois règnes de la nature. Botanique. Paris: Panckoucke.
- Linajes A., Rico-Gray V., Carrion G. 1994. Traditional production system of the root of Jalapa, *Ipomoea purga* (Convolvulaceae), in Central Veracruz, Mexico. *Econ. Bot.* 48:84–89.

- Lowe S.B., Wilson L.A. 1974a. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 2. Interrelationships between tuber shape and yield. *Ann. Bot.* 38:319–326.
- Lowe S.B., Wilson L.A. 1974b. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 1. Tuber initiation, tuber growth and partition of assimilate. *Ann. Bot.* 38:307–317.
- Mabberley D.J. 2008. *Mabberley's Plant Book*. Cambridge University Press.
- Mallet J. 2007. Hybrid speciation. *Nature*. 446:279–283.
- Manos P.S., Miller R.E., Wilkin P. 2001. Phylogenetic Analysis of *Ipomoea*, *Argyreaia*, *Stictocardia*, and *Turbina* Suggests a Generalized Model of Morphological Evolution in Morning Glories. *Syst. Bot.* 26:585–602.
- McDonald J.A. 1989. Neotypification of *Ipomoea jalapa* (Convolvulaceae). *Taxon*. 38:135–138.
- McDonald J.A. 1991. Origin and Diversity of Mexican Convolvulaceae. *An. Inst. Bio.* 62:65–82.
- McDonald J.A. 1994. Convolvulaceae II. In: Sosa V., editor. *Flora de Veracruz*. p. 1–133.
- Mejia-Aguero L.E., Galeno F., Hernandez-Hernandez O., Matehus J., Tovar J. 2012. Starch determination, amylose content and susceptibility to in vitro amylolysis in flours from the roots of 25 cassava varieties. *J. Sci. Food Agric.* 92:673–678.
- Nedunchezhiyan M., Byju G., Jata S.K. 2012. Sweet Potato Agronomy. *Fruit, Veg. Cereal Sci. Biotechnol.* 6:1–10.
- Noh S.A., Lee H.-S., Huh E.J., Huh G.H., Paek K.-H., Shin J.S., Bae J.M. 2010. SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (*Ipomoea batatas*). *J. Exp. Bot.* 61:1337–49.
- Noh S.A., Lee H.-S., Kim Y.-S., Paek K.-H., Shin J.S., Bae J.M. 2013. Down-regulation of the IbEXP1 gene enhanced storage root development in sweetpotato. *J. Exp. Bot.* 64:129–142.
- Otto S.P., Whitton J. 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* 34:401–437.
- Ozias-Akins P., Jarret R.L. 1994. Nuclear DNA Content and Ploidy Levels in the Genus *Ipomoea*. *J. Soc. Hortic. Sci.* 119:110–115.
- Ramsey J., Schemske D.W. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29:467–501.
- Roullier C., Duputié A., Wennekes P., Benoit L., Fernández Bringas V.M., Rossel G., Tay D., McKey D., Lebot V. 2013. Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS One*. 8:e62707.

- Roullier C., Rossel G., Tay D., McKey D., Lebot V. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Mol. Ecol.* 20:3963–77.
- Ruiters C., McKenzie B. 1994. Seasonal Allocation and Efficiency Patterns of Biomass and Resources in the Perennial Geophyte *Sparaxis grandiflora* Subspecies *fimbriata* (Iridaceae) in Lowland Coastal Fynbos, South Africa. *Ann. Bot.* 74:633–646.
- Ruiters C., McKenzie B., Aalbers J., Raitt L.M. 1993. Seasonal allocation of biomass and resources in the geophytic species *Haemanthus pubescens* subspecies *pubescens* in lowland coastal fynbos, South Africa. *South African J. Bot.* 59:251–258.
- Sankaraman S., Mallick S., Dannemann M., Prufer K., Kelso J., Paabo S., Patterson N., Reich D. 2015. The landscape of Neandertal ancestry in present-day humans. *Nature.* 507:354–357.
- Sargent E. 1904. The Evolution of Monocotyledons. *Bot. Gaz.* 37:325–345.
- Schnable J.C., Springer N.M., Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci.* 108:4069–4074.
- Schumer M., Cui R., Powell D.L., Rosenthal G.G., Andolfatto P. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25:2661–2679.
- De Souza J.G., Viera Da Silva J. 1987. Partitioning of carbohydrates in annual and perennial cotton (*Gossypium hirsutum* L.). *J. Exp. Bot.* 38:1211–1218.
- Steiner U., Leibner S., Schardl C.L., Leuchtmann A., Leistner E. 2011. *Periglandula*, a new fungal genus within the Clavicipitaceae and its association with Convolvulaceae. *Mycologia.* 103:1133–1145.
- Tank D.C., Olmstead R.G. 2008. From annuals to perennials: Phylogeny of subtribe Castillejiniae (Orobanchaceae). *Am. J. Bot.* 95:608–625.
- The Heliconius Genome Consortium, Dasmahapatra K.K., Walters J.R., Briscoe A.D., Davey J.W., Whibley A., Nadeau N.J., Zimin A. V., Hughes D.S.T., Ferguson L.C., Martin S.H., Salazar C., Lewis J.J., Adler S., Ahn S.-J., Baker D. a., Baxter S.W., Chamberlain N.L., Chauhan R., Counterman B. a., Dalmay T., Gilbert L.E., Gordon K., Heckel D.G., Hines H.M., Hoff K.J., Holland P.W.H., Jacquín-Joly E., Jiggins F.M., Jones R.T., Kapan D.D., Kersey P., Lamas G., Lawson D., Mapleson D., Maroja L.S., Martin A., Moxon S., Palmer W.J., Papa R., Papanicolaou A., Pauchet Y., Ray D. a., Rosser N., Salzberg S.L., Supple M. a., Surridge A., Tenger-Trolander A., Vogel H., Wilkinson P. a., Wilson D., Yorke J. a., Yuan F., Balmuth A.L., Eland C., Gharbi K., Thomson M., Gibbs R. a., Han Y., Jayaseelan J.C., Kovar C., Mathew T., Muzny D.M., Ogeri F., Pu L.-L., Qu J., Thornton R.L., Worley K.C., Wu Y.-Q., Linares M., Blaxter M.L., French-Constant R.H., Joron M., Kronforst M.R., Mullen S.P., Reed R.D., Scherer S.E., Richards S., Mallet J., Owen McMillan W., Jiggins C.D. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature.* 487:94–98.

- Veselý P., Bureš P., Šmarda P. 2013. Nutrient reserves may allow for genome size increase: Evidence from comparison of geophytes and their sister non-geophytic relatives. *Ann. Bot.* 112:1193–1200.
- Veselý P., Bureš P., Šmarda P., Pavlíček T. 2012. Genome size and DNA base composition of geophytes: The mirror of phenology and ecology? *Ann. Bot.* 109:65–75.
- Wendel J.F. 2015. The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102:1753–1756.
- Wilson L.A., Lowe S.B. 1973. The anatomy of the root system in West Indian sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. *Ann. Bot.* 37:633–643.
- You M.K., Hur C.G., Ahn Y.S., Suh M.C., Jeong B.C., Shin J.S., Bae J.M. 2003. Identification of genes possibly related to storage root induction in sweetpotato. *FEBS Lett.* 536:101–105.

CHAPTER II:  
PHYLOGENETICS AND DIVERSIFICATION OF MORNING GLORIES (TRIBE IPOMOEAE,  
CONVOLVULACEAE) BASED ON WHOLE PLASTOME SEQUENCES.<sup>1</sup>

<sup>1</sup>Eserman LE, Tiley GP, Jarret RL, Leebens-Mack JH, Miller RE. American Journal of Botany. Accepted 09 October 2013. Reprinted here with permission of publisher.

## **Abstract**

*Premise:* Morning glories are an emerging model system, and resolving phylogenetic relationships is critical for understanding their evolution. Phylogenetic studies demonstrated the largest morning glory genus, *Ipomoea*, is not monophyletic, and nine other genera are derived from within *Ipomoea*. Therefore, systematic research is focused on the monophyletic tribe Ipomoeae (ca. 650-900 species). We used whole plastomes to infer relationships across Ipomoeae.

*Methods:* Whole plastomes were sequenced for twenty-nine morning glory species, representing major lineages. Phylogenies were estimated using alignments of eighty-two plastid genes and whole plastomes. Divergence times were estimated using three fossil calibration points. Finally, evolution of root architecture, flower color, and ergot alkaloid presence was examined.

*Key results:* Phylogenies estimated from both datasets had nearly identical topologies. Phylogenetic results are generally consistent with prior phylogenetic hypotheses. Higher-level relationships with weak support in previous studies were recovered here with strong support. Molecular dating analysis suggests a late Eocene divergence time for the Ipomoeae. The two clades within the tribe, *Argyreinae* and *Astripomoeinae*, diversified at similar times. Reconstructed most recent common ancestor of the Ipomoeae had blue flowers, an association with ergot-producing fungi, and either tuberous or fibrous roots.

*Conclusions:* Phylogenetic results provide confidence in relationships among Ipomoeae lineages. Divergence time estimation results provide a temporal context for diversification of morning glories. Ancestral character reconstructions support previous findings that morning glory morphology is evolutionarily labile. Taken together, our study provides strong resolution

of the morning glory phylogeny, which is broadly applicable to the evolution and ecology of these fascinating species.

## **Introduction**

*Ipomoea* L. is the largest genus within Convolvulaceae with approximately 500-650 species (Wilkin, 1999; Mabberley, 2008). Molecular phylogenetic studies have found that the genus *Ipomoea* as traditionally recognized is not monophyletic (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002; Stefanovic et al. 2002). Further, none of the three subgenera within *Ipomoea* (subgenera *Ipomoea*, *Quamoclit*, *Eriospermum*; Austin, 1979, 1980) is monophyletic (McDonald and Mabry 1992; Miller et al. 1999, 2004). Therefore, systematic studies of morning glories focus on the monophyletic tribe Ipomoeae (Stefanovic et al. 2003) consisting of ca. 650-900 species distributed throughout the tropics and subtropics of the world (Wilkin, 1999; Mabberley, 2008). The spiny pollen of species within the Ipomoeae is distinct from the smooth pollen of the sister tribe Merremieae (Hallier 1893; Stefanovic et al. 2002, 2003). *Ipomoea* and nine other genera, i.e. *Argyreia* Lour. (90 species including *Rivea*), *Turbina* Raf. (15 species), *Astripomoea* A. Meeuse (12 species), *Stictocardia* Hallier f. (12 species), *Lepistemon* Blume (10 species), *Rivea* Choisy (4 species), *Blinkworthia* Choisy (2 species), *Lepistemonopsis* Dammer (1 species), and *Paralepistemon* Lejoly & S. Lisowski (1 species), make up the Ipomoeae (Wilkin 1999; Manos et al. 2001; Stefanovic et al. 2003; Mabberley 2008). Stefanovic et al. (2003) divided the Ipomoeae into two major clades, *Astripomoeinae* and *Argyreinae*, based on phylogenetic analyses of four chloroplast loci, however, these lineages have no obvious distinguishing morphological features (Stefanovic et al. 2003). Generally, the *Argyreinae* contains more paleotropical species, while the *Astripomoeinae* has more neotropical species (Stefanovic et al. 2003), although this pattern may be an artifact of

limited sampling among paleotropical species in phylogenetic studies. This group of species is important economically and has served as a model for understanding many evolutionary questions; therefore, understanding evolutionary relationships among these species is significant area of research.

A well-resolved phylogeny of the Ipomoeae is necessary to address many questions concerning the evolutionary history of morning glories. For example, Austin (1997) notes that species that produce tuberous roots are found scattered across the taxa of American *Ipomoea*. From our understanding of phylogenetic relationships for these species (McDonald and Mabry 1992; Miller et al. 1999, 2002, 2004; Manos et al. 2001; McDonald et al. 2011) we can deduce that tuberous roots have been independently derived multiple times in morning glories. Morning glories are generally known to have fibrous roots, e.g. *Ipomoea purpurea* (L.) Roth, *I. nil* (L.) Roth, *I. pes-caprae* (L.) R.Br. However, many species unrelated to sweetpotato (*I. batatas* (L.) Lam.) produce tuberous roots, e.g. *I. carnea* Jacq., *I. lindheimeri* A. Gray, *I. pandurata* (L.) G. Mey. (Austin 1978, 1997; Horak and Wax 1991; McDonald 1994). Furthermore, on a fine scale, there are many closely-related pairs of species where one member has fibrous roots and the other produces tuberous roots, e.g. *I. pubescens* Lam. (tuberous) and *I. purpurea* (fibrous), *I. plummerae* A. Gray (tuberous) and *I. costellata* Torr. (fibrous), *I. purga* (Wender.) Hayne (tuberous) and *I. dumetorum* Willd. ex Roem. & Schult. (fibrous) (McDonald 1994; Manos et al. 2001; Miller et al. 2004). Many tubers are edible, e.g. *I. pandurata*, *M. dissecta* (Jacq.) Hallier f., or are used in medicine for their purgative properties, e.g. *I. jalapa* (L.) Pursh, *I. purga*, *I. orizabensis* (G. Pelletan) Ledeb. ex Steud. (Noda et al. 1987; McDonald 1989; Horak and Wax 1991; Austin 2007).

A second major area of active research in morning glories is the evolution of flower color and the evolutionary genetics of the anthocyanin biosynthetic pathway, which produces red and blue/purple floral pigments (reviewed in both Rausher, 2008; and Wessinger and Rausher, 2012). Examining the floral transitions among various morning glory species has been instrumental in furthering our understanding of the genetic basis of adaptive evolution. Red flowers, for example, have evolved independently at least four times within the Astropomoeinae (species of *Ipomoea* section *Mina* (Cerv.) Griseb. (*Ipomoea quamoclit* L. and *I. coccinea* L.) as well as *I. urbinei* House, *I. conzattii* Greenm., and *I. horsfalliae* Hook.) and once in the Argyreinae (some species of *Stictocardia*) (Austin et al. 1978; Miller et al. 2004; Streisfeld and Rausher 2009). Loss of floral anthocyanins has occurred independently seven times within the Quamoclit group alone (ca. 84 species) (Smith et al. 2010). One emerging pattern is that changes in transcription factors and more generally regulatory regions most commonly lead to adaptive flower color evolution (Streisfeld and Rausher 2009, 2010; Wessinger and Rausher 2012). For example, the transition from blue to red flowers is attributable to regulatory gene action in *I. coccinea*, *I. horsfalliae*, and *I. quamoclit*. To strengthen these conclusions will require enumeration of many cases documenting the molecular genetic basis of flower color transitions. A phylogenetic perspective will be crucial to determine the nature of the transitions (e.g. blue to red) and whether these are independent events.

Some species of morning glories have long been known to contain ergot alkaloids, especially based on assays of seeds (Hofmann 1961, 2006). For grass species, it has been well established that ergot alkaloids are produced in association with endosymbiotic clavicipitaceous fungi (Schardl and Clay 1997; Schardl et al. 2004). Only recently has it been discovered that ergot alkaloid presence in morning glories is the result of a symbiosis with clavicipitaceous fungi

as well (Kucht et al. 2004; Ahimsa-Müller et al. 2007; Steiner et al. 2011). Specifically, only members of the monophyletic tribe Ipomoeae (including members of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina*) have been found to be ergot-positive (Eich 2008). Drawing from a careful survey of studies of ergot alkaloids in morning glories by Eich (2008), we can estimate that approximately 50% of Ipomoeae contain ergot alkaloids. If we assume each morning glory host species harbors a unique fungal symbiont, then there may be as many as 450 species of clavicipitaceous fungi to be discovered. Furthermore, 46 morning glory species considered in Eich's analyses can be confidently placed within the two main clades of Ipomoeae, the Argyreiinae and Astripomoeinae. From this we find that 62% of species in the Argyreiinae clade are ergot positive (8 ergot positive, 5 ergot negative) and 52% of species in the Astripomoeinae clade are ergot positive (17 ergot positive, 16 ergot negative). While these are very modest samples, they do suggest the Argyreiinae clade may contain a concentration of morning glories that are hosts of clavicipitaceous fungi. To date, two fungal species have been characterized and named *Periglandula ipomoeae* U. Steiner, E. Leistner & Schardl and *P. turbinae* U. Steiner, E. Leistner & Schardl after their two respective host species, *I. asarifolia* (Desr.) Roem. & Schult. and *T. corymbosa* (L.) Raf. (Steiner et al. 2011). Therefore, examining the biodiversity and determining the phylogenetic relationships among potential *Periglandula* species, as well as evaluating how the *Periglandula* phylogeny may relate to morning glory evolutionary relationships, or alternatively the biogeography of endosymbionts, are exciting new areas of investigation.

Clearly, a more complete understanding of morning glory species relationships is critical for comparative analyses of interesting morphological, chemical, reproductive and ecological traits. Relationships among major lineages within the Astripomoeinae and Argyreiinae have been

particularly confusing. Therefore, this study attempts to resolve higher-level relationships within the Ipomoeae. Previous phylogenetic studies of the Ipomoeae have utilized morphology (Wilkin 1999), chloroplast RFLPs (McDonald and Mabry 1992) or one to a few loci (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002; Stefanovic et al. 2002). The present study assesses phylogenetic relationships among the major morning glory lineages using whole chloroplast genome sequences for 29 species. In addition, this study attempts to put the diversification of morning glories within a temporal context with a divergence time analysis of the Convolvulaceae including Solanaceae species using 79 chloroplast genes. The evolution of three traits of major importance is evaluated with ancestral character state reconstructions. Finally, the need for a phylogenetic subtribal classification of the Ipomoeae is discussed.

## **Materials and Methods**

### *Taxon sampling*

Species were sampled to represent the major Ipomoeae lineages, as determined from previous phylogenetic analyses of morning glories, nineteen from the Astripomoeinae, eight from the Argyreiinae, and two outgroup species (Table S2.1; Miller et al., 1999, 2002, 2004; Wilkin, 1999; Manos et al., 2001; Huelsenbeck et al., 2002). To aid genome assembly, the sampling was concentrated among species related to *Ipomoea purpurea* (L.) Roth, the previously published chloroplast genome (McNeal et al. 2007), starting with the very closely-related *I. nil* (L.) Roth and then sampling from there in a nested fashion. Representatives from the largest Ipomoeae genera, i.e. *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* were sampled. Accessions of the other Ipomoeae genera were not sampled because they represent only a small portion of diversity in the Argyreiinae clade (Stefanovic et al. 2002, 2003). While the species included in this study encompass a wide range of morphological diversity, this sparse sampling was not

intended to represent the pattern of morphological variation within Ipomoeae, especially given the high degree of evolutionary lability across the tribe (Manos et al. 2001).

A more focused examination of relationships among the various named species of the sweetpotato complex included three samples of *Ipomoea batatas* (L.) Lam. (sweetpotato), two *I. trifida* (Kunth) G. Don individuals, and one sample each of *I. cordatotriloba* Dennst. and *I. splendor-sylvae* House (= *I. umbraticola* House). Two species of the sister tribe Merremieae (sensu Stefanovic et al., 2003), *Merremia quinquefolia* (L.) Hallier f. and *Operculina macrocarpa* (L.) Urb., were chosen as outgroups for phylogenetic analyses. In total, thirty-three individuals representing thirty species were included in all analyses (Tables S2.1, S2.2).

#### *Plastome sequencing and assembly*

DNA was prepared for sequencing in one of two ways: chloroplast enrichment using a sucrose gradient followed by rolling circle amplification (following Jansen et al. 2005) or an extraction of total genomic DNA using the Qiagen DNeasy Plant Kit (Valencia, California, USA). The amount of chloroplast DNA present in each sample was measured by quantitative real-time PCR of a small region of *rbcL*. DNA templates were sequenced using either Illumina or Roche 454 sequencing platforms (Table S2.2). Illumina sequencing was performed at Cold Spring Harbor Labs (GA2 Illumina Sequencer), with a few exceptions. *Ipomoea batatas* (L.) Lam. PI 508520, PI 518474 and PI 561258 and *I. trifida* (Kunth) G. Don PI 618966 were sequenced with paired-end Illumina sequencing at BGI Americas lab in Davis, CA. Roche 454 sequencing was done at the Georgia Genomics Facility at the University of Georgia.

Reads were assembled using the reference-based assembler YASRA (Ratan 2009) and the de novo assembler Velvet (Zerbino and Birney 2008). VelvetOptimiser (<http://bioinformatics.net.au/software.velvetoptimiser.shtml>) was used to determine kmer size for

assembly. The published *Ipomoea purpurea* (L.) Roth chloroplast genome (McNeal et al., 2007; GenBank accession NC\_009808) was used as the reference genome for YASRA assemblies, except in the case of *Argyria nervosa* (Burm. f.) Bojer, where a more closely related species from this study, *I. pes-tigridis* L., was used as the reference sequence. Contigs generated in YASRA and Velvet were merged in Sequencher v5.0. Reads were mapped back to merged assemblies to verify assembly quality. Reads were mapped using either Bowtie for Illumina reads (Langmead et al. 2009) or MOSAIK (<https://code.google.com/p/mosaik-aligner/>) for 454-based reads. Mapped reads were visualized in Geneious v6.0.5. Merged assemblies were manually adjusted to reflect read support. Mapped reads were used to calculate depth of coverage in the Integrative Genomics Viewer (Table S2.2; Robinson et al., 2011; Thorvaldsdóttir et al., 2013). Assembled plastomes were annotated using the DOGMA pipeline (Wyman et al. 2004), which utilizes BLAST and a database of fully-annotated plastomes to identify protein-coding, rRNA, and tRNA genes. DNA sequences for the 82 protein-coding and rRNA genes were extracted from the plastomes using DOGMA's sequence extraction function. Inverted repeat boundaries were identified by performing a nucleotide BLAST (blastn) of a plastome to itself (Table S2.3). Plastome sequences were deposited in GenBank under accession numbers KF242473 to KF242504 (Table S2.1).

#### *Gene sampling and DNA sequence alignment*

Two datasets were generated for phylogenetic analyses. One is a concatenated dataset comprised of 82 protein-coding and rRNA genes from the large single copy, the first inverted repeat, and small single copy regions. The other is a whole plastome alignment, where the second inverted repeat was removed. The whole plastome dataset was aligned using Mauve (Darling et al. 2010) and SATé (Liu et al. 2009, 2012). For the 82-gene dataset, individual genes

were aligned in Muscle (Edgar 2004a, 2004b) and SATé (Liu et al. 2009, 2012), and a Perl script was written to concatenate aligned genes. The plastid genome is a single non-recombining molecule, so all single genes included in the concatenated alignment share the same history (Moore et al. 2010).

### *Phylogenetic analyses*

Maximum parsimony, maximum likelihood and Bayesian analyses were performed on the whole plastome and 82-gene alignments. The most appropriate model of nucleotide substitution for Bayesian and maximum likelihood analyses (GTR+I+ $\Gamma$ ) was inferred using jModelTest2 (Darriba et al. 2012). One substitution model (GTR+I+ $\Gamma$ ) was applied to both datasets. A maximum likelihood bootstrap analysis sampling 500 pseudoreplicates was performed for both the whole plastome and 82-gene datasets using RAxML v7.3.0 (Stamatakis 2006). Bayesian analyses were performed using MrBayes version 3.2.1 (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012). Markov chain Monte Carlo as implemented in MrBayes was conducted using two independent runs and four chains, sampling every 200 generations for a total of 20 million generations. Chains were determined to have converged when 50% majority-rule consensus trees from both independent runs exhibited the same topology, and posterior probabilities of clade support were within a range of 3% (Huelsenbeck et al. 2002). The burn-in fraction was established using a plot of total tree length by generations, a conservative measure of burn-in (Miller et al. 2004). For all analyses, the first 25% of samples were removed as burn-in. A maximum parsimony bootstrap analysis was performed sampling 100 pseudoreplicates using one random taxon addition per replicate in PAUP (Swofford, 2003).

### *Divergence time estimation*

Divergence times were estimated to place the evolution of characters in a temporal context and to understand how the timing of morning glory diversification compares to other angiosperm groups. We applied a Bayesian divergence time analysis in BEAST v1.7.2 (Drummond et al. 2012). We applied an uncorrelated log-normal relaxed clock model, which allows each branch to have its own substitution rate drawn from a log-normal distribution (Drummond et al. 2006). Three Solanaceae species, *Solanum tuberosum*, *Nicotiana tabacum*, and *Atropa belladonna*, were added for divergence time analyses. Solanaceae plastid gene sequences were obtained from the MonATol Plastid gene database (<http://jlmwiki.plantbio.uga.edu/PlastidDB/>). A dataset consisting of 79 chloroplast genes aligned in SATé (Liu et al. 2009, 2012) was used to estimate divergence times. The Yule prior was applied to estimate the branching process. A single model of nucleotide substitution (GTR+I+ $\Gamma$ ) was assumed for the entire dataset.

Three nodes were calibrated with fossil pollen placed within well-defined geological strata (Geological Society of America, 2012). For the BEAST analysis, boundary ages for calibration nodes were set to the youngest epoch age for the geological stratum in which each fossil was preserved. The age of the crown group for Solanaceae species belonging to the “x=12” clade, including the Nicotianoideae and Solanoideae clades, was calibrated using a *Solanum*-like pollen fossil from Oligocene, i.e. 23.0-33.9 million years ago (mya), deposits in Mexico (Martinez-Hernandez and Ramirez-Arriaga, 1999; Graham, 2010). A *Calystegiapollis microechinatus* fossil pollen from the Lower Eocene deposits in Cameroon was used to calibrate the stem group for the Convolvulaceae at 47.8-56.0 mya (Muller, 1981). *Merremia* Dennst. ex Endl. fossil pollen from the middle Eocene deposits in Brazil, Colombia, and Nigeria were used

to calibrate the root of most recent common ancestor of the two *Merremieae* species (*Merremia* and *Operculina* Silva Manso) at 41.2-47.8 mya (Pares Regali et al. 1974a, b; Legoux 1978; Muller 1981). An exponential prior was applied to the three nodes calibrated with fossil pollen data, which assumes the date of the fossil is close to the age of the node being calibrated (Ho and Phillips 2009). An exponential prior was chosen over other calibration priors such as a gamma or log-normal distribution because the Convolvulaceae pollen is well-represented in the fossil record (Graham and Jarzen 1969; Muller 1981; Martin 2000, 2001; Graham 2010). The Solanales, Solanaceae, Convolvulaceae, Ipomoeae and *Merremieae* nodes were constrained to be monophyletic. Markov chain Monte Carlo was continued for 100 million generations, sampling every 1000 generations initiating from a random starting tree. Convergence of two independent runs was determined using Tracer v.1.4 (Rambaut and Drummond 2007), and the burn-in fraction was 25%.

#### *Character evolution*

Character states were obtained from published literature for three characters of broad agricultural and evolutionary interest, i.e. root architecture, flower color, and ergot alkaloid presence. Character states with references can be found in Table S2.4. Ancestral character states were reconstructed for each character using Mesquite v.2.75 (Maddison and Maddison 2011). A likelihood approach using the Mk1 model was applied in Mesquite. The Mk1 model is a modification of the Jukes-Cantor model of DNA substitution and the Mk model of Lewis (2001), where there is an equal probability of switching between discrete character states.

Ancestral character states were reconstructed using the tree topology and branch lengths of the 50% majority-rule consensus tree from the Bayesian analysis of the Mauve-aligned whole plastome dataset. Two *Ipomoea batatas* (L.) Lam. individuals were removed from the tree

because Mesquite treats all terminal taxa as separate species. Therefore, having multiple individuals with the same character state can overly influence the ancestral reconstructions. Both *I. trifida* (Kunth) G. Don accessions were retained for ancestral character state reconstructions because they did not form a monophyletic species. Taxa with missing data were treated as missing from the tree.

## Results

### *Chloroplast genome structure*

The thirty-two plastomes sequenced and the published *Ipomoea purpurea* (L.) Roth plastome were completely collinear (Figure 2.1 shows *I. hederifolia* L.). Whole plastome sequences ranged from 159848 to 162850 nucleotides long, and GC content was 37% for all plastomes (Table S2.3). Inverted repeat boundaries were generally consistent among species (Table S2.3). The boundary between the large single copy region (LSC) and one inverted repeat (IR<sub>A</sub>) was between *rpl23* and *trnI-CAU* for most analyzed plastomes. However, the LSC-IR<sub>A</sub> boundary was between *trnI-CAU* and *ycf2* in *Stictocardia macalusoii* (Mattei) Verdc. and in *ycf2* in *I. involucrata* P. Beauv. and *Argyreia nervosa* (Burm. f.) Bojer. The boundary between the IR<sub>A</sub> and the small single copy region (SSC) was between *ndhH* and *ndhF* in all species. The SSC-IR<sub>B</sub> boundary was in exon 1 of *ndhA* for all species except *I. pes-tigridis* L. where the SSC-IR<sub>B</sub> boundary was in the *ndhA* intron. Finally, the IR<sub>B</sub>-LSC boundary was between *trnH-GUG* and *trnI-CAU* in all species.

### *Data matrices*

After the second inverted repeat was removed from all sequences, the whole plastome SATé alignment was 140496 bp long, and the whole plastome Mauve alignment was 140818 bp long. The 82-gene alignment was 74315 bp long from SATé and 74262 bp long from Muscle.

Both whole plastome alignments had 3% parsimony informative sites (>4000 sites), and both concatenated 82-gene alignments had 2% parsimony informative sites (>1400 sites), indicating that for these taxa chloroplast sequences are generally conserved, but the majority of informative sites in the plastome lie in intergenic regions.

### *Phylogenetic analyses*

Tree topologies were generally consistent across all phylogenetic analyses for all datasets. The only different topology was recovered in the parsimony tree of the mauve-aligned whole plastome dataset, where there was weak support for *Ipomoea batatas* (L.) Lam. PI 561258 and *I. trifida* (Kunth) G. Don as sister to one another (BS=54). In all other analyses, accessions of *I. batatas* formed a well-supported monophyletic group. ML bootstrap support was generally lower in the 82-gene phylogeny compared to the whole plastome phylogeny. Low likelihood and parsimony bootstrap support values tended to fall on short branches (Figures 2.2, S2.2). There was support for the monophyly of the tribe Ipomoeae and the two major clades, Argyreiinae and Astripomoeinae in all analyses (BS=100; PP=1.0). Within the Astripomoeinae, two major clades were recovered, the Cairica clade and the larger clade consisting of the Batatas, Mururoides, Pes-caprae, and Quamoclit clades. Within the larger Astripomoeinae clade, four smaller clades were recovered, i.e. the Batatas, Mururoides, Pes-caprae, and Quamoclit clades. Within the Argyreiinae, two major clades were recovered, i.e. the Pes-tigridis and Obscura clades. Lowest support values were observed for the Cairica clade (BS=52-82; PP=0.98-1.0).

### *Divergence time estimation*

Mean age of the common ancestor of the Ipomoeae is ca. 35 my (Table 2.1, Figure 2.3). In addition, these results suggest the Argyreiinae and Astripomoeinae clades diverged around the

same time period (Table 2.1, Figure 2.3). The Murucoides clade was the youngest named clade, having diversified ca. 5 mya (Table 2.1, Figure 2.3).

### *Character evolution*

Figures depicting likelihood-based ancestral character state reconstructions can be found in Figure S2.1. For root architecture, most ancestral nodes have an equal probability of having either fibrous or tuberous roots. Therefore, there are either ten independent origins of tubers in *Ipomoea argillicola* R.W. Johnson, *I. batatas* (L.) Lam. + *I. trifida* (Kunth) G. Don REM 753, *I. cairica* (L.) Sweet, *I. dumetorum* Willd. ex Roem. & Schult., *I. orizabensis* (G. Pelletan) Ledeb. ex Steud., *I. pedicellaris* Benth., *I. polpha* R.W. Johnson, *I. setosa* Ker Gawl., *I. ternifolia* Cav., and *I. trifida* PI 618966. Alternatively, tubers were lost independently ten times in *I. amnicola* Morong, *I. cordatotriloba* Dennst., *I. hederifolia* L., *I. involucrata* P. Beauv. + *I. pes-tigridis* L., *I. minutiflora* (M. Martens & Galeotti) House, *I. murucoides* Roem. & Schult., *I. nil* (L.) Roth + *I. purpurea* (L.) Roth, *I. obscura* (L.) Ker Gawl., *I. pes-caprae* (L.) R. Br., and *I. tricolor* Cav. In most instances, the most likely ancestral flower color across the Ipomoeae was blue/purple flowers. There were six evolutionary transitions to white flowers in *I. diamantinensis* J.M. Black, *I. minutiflora*, *I. murucoides*, *I. obscura*, *I. pes-tigridis*, and *Turbina corymbosa* (L.) Raf. Furthermore, there were two transitions from blue/purple flowers to red flowers in *I. hederifolia* and *Stictocardia macalusoii* (Mattei) Verdc. With respect to ergot alkaloid presence, the ancestor of the Ipomoeae was ergot positive (contained ergot-producing fungi) and there were four subsequent losses of ergot-producing endosymbionts in the Batatas + Murucoides clade, *I. eriocarpa* R. Br. + *I. involucrata*, *I. hederifolia* + *I. ternifolia*, and *I. obscura*.

## Discussion

### *Phylogenetic relationships*

Results of this study support the monophyly of the tribe Ipomoeae and its two major clades, the Astripomoeinae and Argyreiinae (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002; Stefanovic et al. 2002). The tribe Ipomoeae was recently expanded by Stefanovic et al. (2003) to include all morning glory species with spiny pollen, uniting the Argyreieae and Ipomoeae tribes proposed by Hallier (1893). This expanded Ipomoeae is consistent with Hallier's subfamily Echinoconiae and encompasses *Ipomoea* and nine other genera (*Argyreia*, *Astripomoea*, *Blinkworthia*, *Lepistemon*, *Lepistemonopsis*, *Paralepistemon*, *Rivea*, *Stictocardia* and *Turbina*) (Manos et al. 2001; Stefanovic et al. 2003). In contrast, no clear morphological features distinguish the Astripomoeinae and Argyreiinae clades (Miller et al. 2002; Stefanovic et al. 2003). The Astripomoeinae primarily consists of New World species, while the Argyreiinae consists of mostly Old World species (Stefanovic et al. 2003). However, there are many exceptions to this pattern, e.g. neotropical *I. pedicellaris* Benth., *Turbina cordata* (Choisy) Austin and Stapes, and *Turbina corymbosa* (L.) Raf. are members of the Argyreiinae and the Australian endemics *I. argillicola* R.W. Johnson, *I. polpha* R.W. Johnson, and *I. diamantinensis* J.M. Black, as well as the Asian *I. sumatrana* (Miq.) Ooststr. are in the Astripomoeinae.

Within the Argyreiinae and Astripomoeinae, several smaller clades were recovered, and relationships among these clades are strongly supported (Figures 2.2, S2.2). Many of these clades were recovered with varying degrees of support in prior phylogenetic investigations, but relationships among these clades were not clear. Well-supported major lineages are given provisional clade names here based on the oldest species within the clade included in this study.

## Astripomoeinae

Within the Astripomoeinae, relationships among the major lineages were well-resolved (Figures 2.2, S2.2). The present analysis recovered the Batatas and Mururoides clades as sister to one another. The Batatas and Mururoides clade was recovered as sister to the Pes-caprae clade. The larger clade containing the Batatas, Mururoides, and Pes-caprae groups was then recovered as sister to the Quamoclit clade. The relationship between the Batatas and Mururoides clades was the most well-supported in other phylogenetic analyses of morning glories (Huelsenbeck et al. 2002; Miller et al. 2002; McDonald et al. 2011). The Batatas, Mururoides, Pes-caprae, and Quamoclit clades were recovered as most closely related to one another with strong support in most other systematic studies of morning glories (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002; McDonald et al. 2011). The Cairica clade was the most basal member of the Astripomoeinae clade in this analysis; however, phylogenies of morning glories using ITS and waxy have typically found the Cairica clade to be sister to the Quamoclit group (Miller et al. 1999, 2002; Manos et al. 2001; Huelsenbeck et al. 2002).

The Quamoclit group consists of approximately 84 neotropical species and is one of the most intensively studied groups of morning glories to date (Miller et al. 2004; Smith et al. 2010; McDonald et al. 2011). Species within the Quamoclit clade have been studied as a model for understanding flower color evolution and the molecular genetics of the anthocyanin biosynthetic pathway (Clegg and Durbin, 2003; Rausher, 2008; Baucom et al., 2011; Wessinger and Rausher, 2012). Quamoclit species exhibit a wide range of pollination syndromes from bee to hummingbird and hawkmoth (McDonald 1991; Miller et al. 2004; Smith et al. 2010). Support for the monophyly of this group had been previously established (Miller et al. 1999; Manos et al. 2001; Huelsenbeck et al. 2002; McDonald et al. 2011). Interestingly, two clades recovered in

previous phylogenetic analyses of the Quamoclit group (delineated as Clade 1 and Clade 2 by Miller et al., 2004) were not found here. Rather, *Ipomoea nil* (L.) Roth and *I. purpurea* (L.) Roth grouped with *I. hederifolia* L., *I. ternifolia* Cav., and *I. minutiflora* (M. Martens & Galeotti) House rather than *I. tricolor* Cav. and *I. orizabensis* (G. Pelletan) Ledeb. ex Steud. as previously hypothesized (Miller et al. 1999, 2004).

The Batatas clade was monophyletic with strong support. This clade unites species of the Batatas complex with *Ipomoea* section *Setosae* (House) D. F. Austin. The Batatas complex consists of fourteen named species (Austin 1978, 1988a; McDonald and Austin 1990). Previous phylogenetic analyses found strong support for this clade to include species of the Batatas complex, *I. setosa* Ker Gawl. and *I. sepacuitensis* Donn. Sm. (Miller et al. 1999, 2002; Manos et al. 2001; Huelsenbeck et al. 2002; McDonald et al. 2011). The most commercially important species of this group is sweetpotato, *I. batatas* (L.) Lam. Cultivated sweetpotato is a hexaploid, and many other members of the Batatas complex vary in ploidal level, i.e. diploid *I. cordatotriloba*, tetraploid *I. trifida* (Ozias-Akins and Jarret 1994). Taxonomy and species delimitation in the Batatas complex has been particularly difficult because individuals often exhibit intermediate morphologies between descriptions of named species (Austin 1978; McDonald and Austin 1990). Furthermore, many members of the Batatas complex are known to hybridize readily (Diaz et al. 1996). The complexities inherent in the Batatas complex are illustrated in these results with the placement of *I. trifida* (Kunth) G. Don. The two specimens identified as *I. trifida* for this analysis were not recovered as monophyletic. In fact, one *I. trifida* individual grouped with *I. batatas* individuals, and the second *I. trifida* grouped with *I. cordatotriloba* Dennst.

The Murucoides clade consists of *Ipomoea murucoides* Roem. & Schult. and *I. polpha* R.W. Johnson in this analysis. Previous phylogenetic analyses of the Ipomoeae have found strong support for this clade to include species with vastly different morphologies and biogeographic affinities (Miller et al. 1999, 2002; Manos et al. 2001; Huelsenbeck et al. 2002; McDonald et al. 2011). Species of this clade are ground trailing vines (*I. polpha*), erect shrubs (*I. carnea* Jacq., *I. cuneifolia* Meisn.), and trees (*I. murucoides*, *I. pauciflora* M. Martens & Galeotti). Furthermore, these species range from neotropical (*I. murucoides*, *I. carnea*) to Australian endemics (*I. polpha*, *I. costata* F. Muell. ex Benth.) and Asian species (*I. sumatrana* (Miq.) Ooststr.).

Species of the Pes-caprae group exhibit variable morphologies and biogeographic patterns. Species of this group range from Australian endemics, e.g. *Ipomoea argillicola* R.W. Johnson, *I. gracilis* R. Br., *I. muelleri* Benth.; tuber-producing twining vines endemic to the United States, *I. leptophylla* Torr., *I. pandurata* (L.) G. Mey.; and neotropical twining vines, e.g. *I. amnicola* Morong. (Miller et al. 1999, 2002; Manos et al. 2001; Huelsenbeck et al. 2002; McDonald et al. 2011). Interestingly, the Pes-caprae clade is united by their shared association with clavicipitaceous fungal endophytes, which produce ergot alkaloids (Eich 2008).

The Cairica group was sister to the rest of the Astripomoeinae clade and received the lowest support of all clades recovered in this analysis. The phylogenetic affinity of this clade has been uncertain. The Cairica clade is typically sister to the Quamoclit group, but bootstrap support for this topology was always <70% (Miller et al. 1999, 2002; Manos et al. 2001; Huelsenbeck et al. 2002).

## Argyreinae

Within the Argyreinae, two clades were recovered with strong support, the Pes-tigridis and Obscura clades. These two clades were often recovered as sister to one another (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002) and have both been shown to be monophyletic (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002). The Pes-tigridis clade joins species from six different genera, i.e. *Ipomoea*, *Argyreia*, *Lepistemon*, *Stictocardia*, *Rivea*, and *Turbina*, while members of three different genera comprise the Obscura clade, i.e. *Ipomoea*, *Stictocardia*, and *Turbina* (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002). These results coupled with prior phylogenetic hypotheses suggest the genera *Ipomoea* and *Turbina* are not monophyletic (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002). Furthermore, the genus *Stictocardia* is monophyletic but has been recovered in both the Pes-tigridis and Obscura clades (Manos et al. 2001; Huelsenbeck et al. 2002; Miller et al. 2002). Species of the Pes-tigridis clade are typically paleotropical (e.g. *I. pes-tigridis* L., *I. involucrata* P. Beauv., *Argyreia nervosa* (Burm. f.) Bojer). In contrast, the Obscura clade contains both paleotropical (*I. obscura* (L.) Ker Gawl.) and neotropical (*I. pedicellaris* Benth., *T. corymbosa* (L.) Raf.) members.

## *Divergence time estimates*

The divergence time estimated for the Convolvulaceae in this study (47.80-69.98 mya) was similar to the age of the Convolvulaceae estimated in two previous studies. One study estimated the divergence age of the Convolvulaceae to be ca. 50-85 mya (Dillon et al. 2009), and a second study estimated the Convolvulaceae to be ca. 65-66 my (Wikström et al. 2001). The similarity among these estimates is gratifying given that all three studies used different approaches to estimate divergence times. Dillon et al. (2009) examined divergence times in the

Convolvulaceae and Solanaceae, primarily focusing on the genus *Nolana* L. f., using a Bayesian divergence time analysis. The *Nolana* study calibrated three nodes, two using fossils and one using an estimated divergence time between the Solanaceae and Convolvulaceae from Bremer et al. (2004), and a normally distributed prior was applied to the calibrated nodes (Dillon et al. 2009). Wikström et al. (2001) applied a parsimony-based divergence time approach using accelerated and delayed transformation to estimate divergence times across the angiosperms.

Our divergence time estimates show that the Argyreiinae and Astripomoeinae clades have been on separate evolutionary trajectories for ca. 25 million years. However, there are no clear morphological distinctions between the two major clades (Wilkin 1999; Stefanovic et al. 2003). In fact, members of both the Astripomoeinae and Argyreiinae have been placed in the same section by various taxonomists. For example, *Ipomoea purpurea* (L.) Roth and *I. pes-tigridis* L. were in two separate treatments considered members of *Ipomoea* section *Pharbitis* (Hallier 1893; Roberty 1952). These two species exhibit very similar gross morphologies, i.e. annual, weedy habit, herbaceous sepals, dense, hispid trichomes; however, *I. purpurea* belongs to the Astripomoeinae and *I. pes-tigridis* belongs to the Argyreiinae.

The long divergence time between the Astripomoeinae and Argyreiinae is surprising in light of other plant groups estimated to be of a similar age. The MRCA of the Araceae genera *Arum*, *Dracunculus* and *Biarum* is ca. 20-40 (Mansion et al. 2008), and *Arum* + *Dracunculus* form a monophyletic group sister to the genus *Biarum* (Cabrera et al. 2008; Mansion et al. 2008). This split is similar in age to the divergence between Astripomoeinae and Argyreiinae; however, there are morphological distinctions between these clades, i.e. *Dracunculus* and *Arum* have similar flower structures that are distinct from *Biarum* (Boyce 2008). Similarly, in the Solanaceae, the most recent common ancestor to the Solanoideae and Nicotianoideae subfamilies

is ca. 15-25 Ma (Dillon et al. 2009). These two subfamilies make up a monophyletic group referred to as the “x=12” clade (Olmstead et al. 2008). They are distinct from each other in that members of the Nicotianoideae have capsular fruits, while members of the Solanoideae have berries (Knapp 2002). Based on studies of other plant taxa that have diverged over a similar time period we would expect to observe morphological differences between the Argyreiinae and Atripomoeinae. However, no differences have been detected even after tens of millions of years of divergence. The suite of morphological forms and pattern of morphological variation among the two distinct lineages not only is consistent with the high degree of evolutionary lability for morning glories in general but also suggests that parallel evolution may be occurring for these two major clades within the tribe Ipomoeae.

#### *Implications for character evolution*

Previous studies have demonstrated that there is a high degree of evolutionary lability in morning glory morphology, e.g. flower color, locule number, sepal shape (Wilkin 1999; Manos et al. 2001; Smith et al. 2010). Therefore, it is difficult to make specific conclusions regarding character evolution within the Ipomoeae. However, some important patterns of character evolution did emerge.

Our results suggest there are two equally parsimonious scenarios for the evolution of tubers in the Ipomoeae. Therefore, there could have been either ten independent origins or losses of tubers. Little is known about root morphology, especially in Argyreiinae species. With a more complete sampling of species and a more complete phylogeny, this pattern may become better resolved. It would be interesting to see if similar genetic mechanisms are responsible for tuber production in the tuber-producing species across the phylogeny. Root morphology has been used in the subgeneric classification of *Ipomoea* with mixed success. *Ipomoea* section

*Exogonium* (Choisy) Griseb. was defined partly on the basis of having tuberous roots (Austin 1997), but this section is not monophyletic (Miller et al. 2004; McDonald et al. 2011). However, monophyletic *Ipomoea* series *Arborescentes* (Choisy) D.F. Austin and sect. *Mina* (Cerv.) Griseb. are diagnosable partly on the basis of fibrous roots (Austin 1997; Manos et al. 2001; Miller et al. 2004; McDonald et al. 2011).

The ancestral character state reconstructions indicate there were multiple independent flower color transitions. Our results mirror patterns of flower color transitions in other groups of species, e.g. *Antirrhinum* (Jones and Reithel 2001), the Solanaceae clade *Iochrominae* (Smith and Baum 2007), *Mimulus* (Streisfeld and Kohn 2005), and *Ruellia* (Tripp and Manos 2008), supporting the idea that flower color is a highly labile trait. Blue or purple flowers was recovered as the ancestral condition in the Ipomoeae (Figure S2.1), which has previously been hypothesized (McDonald 1991; Smith et al. 2010). In the current phylogeny, there were six losses of anthocyanins in association with the origins of *I. diamantinensis* J. M. Black, *I. minutiflora* (M. Martens & Galeotti) House, *I. murucoides* Roem. & Schult., *I. obscura* (L.) Ker Gawl, *I. pes-tigridis* L., and *Turbina corymbosa* (L.) Raf. In addition, red flowers arose two times independently on this phylogeny in *I. hederifolia* L. and *Stictocardia macalusoii* (Mattei) Verdc.

These results provide excellent additional focal taxa for examining the genetic basis of adaptive evolution by determining the exact mechanisms leading to these flower color transitions. Zufall and Rausher (2004) demonstrated that the transition from blue/purple flowers (in *Ipomoea purpurea* (L.) Roth) to red flowers (in *I. hederifolia* L.) was from a drastic reduction in production of the anthocyanin biosynthetic pathway gene F3'H, likely due to regulatory gene action. Building upon these results, Streisfeld and Rausher (2009) showed that decreased F3'H

expression was responsible for the independent origins of red flowers in the Astripomoeinae. Particularly intriguing would be comparing the genetic basis of red flowers for the members of Astripomoeinae (*I. hederifolia*) that represent the hummingbird pollination syndrome to the African, red-flowered *Stictocardia macalusoii* (Mattei) Verdc. from the Argyreiinae clade suggested to be bird pollinated (Austin and Demissew 1997), these taxa representing lineages that we now know diverged from each other c. 25 mya.

The presence of ergoline alkaloid producing fungal endophytes has been derived multiple times throughout the evolutionary history of morning glories (reviewed in Eich, 2008). Based on the records provided by Eich (2008) and our ancestral character reconstruction (Figure S2.1), we find that having an association with ergot alkaloid-producing fungi is the ancestral condition in the Ipomoeae. Our results indicate ergot fungi have been lost four times. Ergot alkaloids are produced by Clavicipitaceous fungi, which are vertically transmitted via seeds (Schardl et al. 2013). Perhaps these four lineages have lost the ability to vertically transmit the endosymbiotic fungi. This trait has only relatively recently been characterized and has been measured in only a small number of Ipomoeae species. With further sampling, the pattern of ergot alkaloid presence in morning glories can be further illuminated.

All examined species within the Pes-caprae clade, i.e. *Ipomoea amnicola* Morong, *I. argillicola* R.W. Johnson, *I. asarifolia* (Desr.) Roem. & Schult., *I. leptophylla* Torr., *I. muelleri* Benth., and *I. pes-caprae* (L.) R. Br., have been identified as ergot positive. Within the Quamoclit clade, all examined members of *I. sect. Tricolores* J.A. McDonald, i.e. *I. cardiophylla* A. Gray, *I. marginisepala* O'Donell, and *I. tricolor* Cav., have been found to harbor ergot alkaloids (Eich 2008). Additionally, in this analysis *I. tricolor* and *I. orizabensis* (G. Pelletan) Ledeb. ex Steud. were recovered as most closely related to each other, and they are both ergot

positive. Members of *I.* sect. *Mina* (Cerv.) Griseb. generally have conflicting reports of ergot presence (Eich 2008). Sister to the ergot-negative Batatas clade is the Murucoides clade, which has both ergot positive and negative species, i.e. *I. arborescens* (Humb. & Bonpl. ex Willd.) G. Don and *I. murucoides* Roem. & Schult. (ergot negative) and *I. costata* F. Muell. ex Benth. (ergot positive). The Argyreiinae clade appears to be particularly rich in ergot-positive species. For example, all 14 species in the genus *Argyreia* assayed for alkaloids were ergot positive (Eich, 2008).

### *Biogeography*

The biogeographic distribution of many morning glory species has been altered by anthropogenic dispersal, complicating any assessment of broad biogeographic patterns. Some species are ancestrally neotropical but have likely been dispersed by humans to the paleotropics, e.g. *Ipomoea triloba* L., *I. hederifolia* L., *I. nil* (L.) Roth (van Ooststroom and Leyden, 1953; Verdcourt, 1963; Austin et al., 2001). Conversely, other species with a current pantropical distribution are thought to be of paleotropical origins, e.g. *I. cairica* (L.) Sweet, *I. argillicola* R.W. Johnson, *I. pes-caprae* (L.) R. Br. (Shinners 1979; Austin 2005). Many of these species were distributed for use as either medicine or ornamentals (Austin 2000; Austin et al. 2001). However, others are thought to have dispersed by ocean currents (McDonald 1991). In addition, presumably two other main sources of anthropogenic dispersal are in ship ballasts, as well as weed contaminants of crop seeds. Clearly, biogeographic hypotheses need to be explicitly tested using a combined historical, population genetic, and phylogeographic approach (e.g. Richards et al., 2007; Ree and Smith, 2008) as has been done for sweetpotato (Montenegro et al. 2008; Roullier et al. 2011, 2013a, 2013c).

### *Toward a phylogenetic classification of the Ipomoeae*

Morning glory taxonomy and phylogenetic hypotheses for species relationships often are not in congruence (Miller et al. 1999, 2004; Manos et al. 2001; Stefanovic et al. 2002; McDonald et al. 2011). At a broad level, the development of taxonomic classifications for morning glories have been dynamic (e.g. Choisy, 1833; Grisebach, 1864; Hallier, 1893; House, 1908; van Ooststroom and Leyden, 1953; Verdcourt, 1957; Austin, 1975, 1979, 1997; McDonald, 1991), likely owing to the fact that morning glory species are united by rather inconspicuous characters (e.g. spiny pollen of the Ipomoeae, three-locular capsule of *Ipomoea* section *Pharbitis* (Choisy) Griseb., sepals with large dorsal arista of *I. sect. Mina* (Cerv.) Griseb.). In contrast, phylogenetic results based on morphology, DNA sequences and RFLPs are largely congruent (e.g. McDonald and Mabry, 1992; Wilkin, 1999; Manos et al., 2001; Stefanovic et al., 2002; Miller et al., 2004). Therefore, classification of the Ipomoeae based on phylogeny is warranted.

Currently, tribal level taxonomy within the Convolvulaceae based on phylogenetic results delineated the Ipomoeae to represent what was previously the Ipomoeae, plus the Argyreieae (Stefanovic et al. 2003). In addition, the major clades Argyreiinae and Astripomoeinae were defined (Stefanovic et al. 2003). We present clade names to extend this classification further. Clade names presented here identify lineages within the Argyreiinae and Astripomoeinae derived from the most informed hypothesis of evolutionary relationships based on results from multiple phylogenetic studies. However, given that phylogenetic studies to date have sampled ca. 15% of Ipomoeae diversity, a formal classification is not being proposed. Clade names are presented as our best estimate of the sub-tribal groupings within the Ipomoeae.

## Conclusion

Taken together, results of this study have provided a better understanding of the evolutionary relationships and divergence times among morning glory species and established a foundation for future studies. The phylogenomic inferences presented here support the conclusions of prior studies, indicating the need for taxonomic revision of the Ipomoeae and verifying that complex morphological characters, including those of economic importance, have been evolutionarily labile across the morning glory phylogeny. For example, multiple transitions in flower color, in root architecture (fibrous vs. tuberous) and in association with ergot alkaloid producing symbionts are inferred through ML modeling of character evolution.

## References

- Ahimsa-Müller, M. A., A. Markert, S. Hellwig, V. Knoop, U. Steiner, C. Drewke, and E. Leistner. 2007. Clavicipitaceous fungi associated with ergoline alkaloid-containing Convolvulaceae. *Journal of Natural Products* 70: 1955–60.
- Austin, D. F. 1975. Typification of the New World subdivisions of *Ipomoea* L. (Convolvulaceae). *Taxon* 24: 107–110.
- Austin, D. F. 1978. The *Ipomoea batatas* complex. I. Taxonomy. *Bulletin of the Torrey Botanical Club* 105: 114–129.
- Austin, D. F. 1979. An infrageneric classification for *Ipomoea* (Convolvulaceae). *Taxon* 28: 359–361.
- Austin, D. F. 1980. Additional comments on infrageneric taxa in *Ipomoea* (Convolvulaceae). *Taxon* 29: 501–502.
- Austin, D. F. 1988. Nomenclatural Changes in the *Ipomoea batatas* Complex (Convolvulaceae). *Taxon* 37: 184–185.
- Austin, D. F. 1997. Dissolution of *Ipomoea* series *Anisomerae* (Convolvulaceae). *Journal of the Torrey Botanical Society* 124: 140–159.
- Austin, D. F. 2000. The search for “kaladana” (*Ipomoea*, Convolvulaceae). *Economic Botany* 54: 114–118.
- Austin, D. F. 2005. The enigmatic “salsa da rua” - *Ipomoea asarifolia* (Convolvulaceae). *Ethnobotany* 17: 41–48.

- Austin, D. F. 2007. *Merremia dissecta* (Convolvulaceae): Condiment, medicine, ornamental, and weed—A review. *Economic Botany* 61: 109–120.
- Austin, D.F., and S. Demissew. 1997. Unique fruits and generic status of *Stictocardia* (Convolvulaceae). *Kew Bulletin* 52: 161–169.
- Austin, D.F., K. Kitajima, Y. Yoneda, and L. Qian. 2001. A Putative Tropical American Plant, *Ipomoea nil* (Convolvulaceae), in Pre-Columbian Japanese Art. *Economic Botany* 55: 515–527.
- Austin, D. F., D. A. Powell, and D. H. Nicolson. 1978. *Stictocardia tiliifolia* (Convolvulaceae) re-evaluated. *Brittonia* 30: 195–198.
- Baucom, R. S., S. M. Chang, J. M. Kniskern, M. D. Rausher, and J. R. Stinchcombe. 2011. Morning glory as a powerful model in ecological genomics: tracing adaptation through both natural and artificial selection. *Heredity* 107: 377–385.
- Boyce, P. 2008. A taxonomic revision of *Biarum*. *Curtis's Botanical Magazine* 25: 2-17.
- Bremer, K, E. Friis, B. Bremer. 2004. Molecular phylogenetic dating of Asterid flowering plants shows early Cretaceous diversification. *Systematic Biology* 53: 496-505.
- Cabrera, L. I., G. A. Salazar, M. W. Chase, S. J. Mayo, J. Bogner, and P. Dávila. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *American Journal of Botany* 95: 1153-1165.
- Choisy, J. D. 1833. Convolvulaceas orientales. *Memoires de la Societe de Physique et d'Historie Naturelle de Geneve* 6: 430–431.
- Clegg, M. T., and M. L. Durbin. 2003. Tracing floral adaptations from ecology to molecules. *Nature Reviews Genetics* 4: 206-215.
- Darling, A. E., B. Mau, and N. T. Perna. 2010. Progressivemaue: multiple genome alignment with gene gain, loss and rearrangement. *Plos One* 5: e11147.
- Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. Jmodeltest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.
- Diaz, J., P. Schmiediche, and D. F. Austin. 1996. Polygon of crossability between eleven species of *Ipomoea* section *Batatas* (Convolvulaceae). *Euphytica* 88: 189–200.
- Dillon, M. O., T. Tu, L. Xie, V. Quipuscoa Silvestre, and J. Wen. 2009. Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *Journal of Systematics and Evolution* 47: 457–476.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *Plos Biology* 4: 699–710.

- Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with beauti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.
- Edgar, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Edgar, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
- Eich, E. 2008. Solanaceae and Convolvulaceae Secondary Metabolites: Biosynthesis, Chemotaxonomy, Biological and Economic Significance: a handbook. Springer, Berlin.
- Geological Society of America. 2012. Geologic Time Scale. <<http://www.geosociety.org/science/timescale/>>.
- Graham, A. 2010. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments. Missouri Botanical Garden Press, St. Louis.
- Graham, A., and D. M. Jarzen. 1969. Studies in Neotropical Paleobotany. I. The Oligocene Communities of Puerto Rico. *Annals of the Missouri Botanical Garden* 56: 308–357.
- Grisebach, A. H. 1864. Convolvulaceae. In *Flora of the British West Indian Islands*, 466–475. London.
- Hallier, H. 1893. Versuch einer natuerlichen Gliederung der Convolvulaceen auf morphologischer und anatomischer Grundlage. *Botanische Jahrbuecher fuer Systematik* 16: 453–591.
- Ho, S. Y. W., and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* 58: 367–380.
- Hofmann, A. 1961. Die Wirkstoffe der mexikanischen Zauberdroge “Ololiuqui”. *Planta Medica* 9: 354–367.
- Hofmann, A. 2006. LSD-mein Sorgenkind, die Entdeckung einer Wunderdroge. 12th ed. Deutscher Taschenbuch Verlag, München.
- Horak, M. J., and L. M. Wax. 1991. Growth and development of bigroot morningglory (*Ipomoea pandurata*). *Weed Technology* 5: 805–810.
- House, H. D. 1908. The North American species of the genus *Ipomoea*. *Annals of the New York Academy of Sciences* 18: 181–263.
- Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51: 673–688.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

- Jansen, R. K., L. A. Raubeson, J. L. Boore, C. W. DePamphilis, T. W. Chumley, R. C. Haberle, S. K. Wyman, et al. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology* 395: 348–384.
- Jones, K.N., and J.S. Reithel. 2001. Pollinator-mediated selection on a flower color polymorphism in experimental populations of *Antirrhinum* (Schrophulariaceae). *American Journal of Botany* 88: 447–454.
- Knapp, S. 2002. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Journal of Experimental Botany* 53: 2001–2022.
- Kucht, S., J. Gross, Y. Hussein, T. Grothe, U. Keller, S. Basar, W. A. König, et al. 2004. Elimination of ergoline alkaloids following treatment of *Ipomoea asarifolia* (Convolvulaceae) with fungicides. *Planta* 219: 619–25.
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
- Legoux, O. 1978. Quelques especes de pollen caracteristiques du Neogene de Nigeria. *Bulletin des Centres de Recherches Exploration-Production Elf-Aquitaine* 2: 265–317.
- Lewis, P. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913–925.
- Liu, K., S. Raghavan, S. Nelesen, C. R. Linder, and T. J. Warnow. 2009. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* 324: 1561–1564.
- Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder. 2012. Sate-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology* 61: 90–106.
- Lohse, M., O. Drechsel, and R. Bock. 2007. Organellargenomedraw (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* 52: 267-274.
- Lohse, M., O. Drechsel, S. Kahlau, and R. Bock. 2013. Organellargenomedraw—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 10.1093/nar/gkt289.
- Mabberley, D. 2008. *Mabberley's Plant Book*. 3rd ed. Cambridge University Press, Cambridge.
- Maddison, W.P, and D.R. Maddison. 2011. *Mesquite: a modular system for evolutionary analysis, version 2.75*. Computer program and documentation distributed by the authors, website: <http://mesquiteproject.org> [accessed 22 March 2013].
- Mansion, G, G. Rosenbaum, N. Schoenenberger, G. Bacchetta, J. A. Rosselló, and E. Conti. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential

- invasions of the Mediterranean Basin by the angiosperm family Araceae. *Systematic Biology* 57: 269–285.
- Manos, P. S., R. E. Miller, and P. Wilkin. 2001. Phylogenetic analysis of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *Systematic Botany* 26: 585–602.
- Martin, H. A. 2000. Re-assignment of the affinities of the fossil pollen type *Tricolpites trioblatus* Mildenhall and Pocknall to *Wilsonia* (Convolvulaceae) and a reassessment of the ecological interpretations. *Review of Paleobotany and Palynology* 111: 237–251.
- Martin, H. A. 2001. The family Convolvulaceae in the Tertiary of Australia: evidence from pollen. *Australian Journal of Botany* 49: 221–234.
- Martinez-Hernandez, E., and E. Ramirez-Arriaga. 1999. Palinostratigrafia de la region de Tepexi de Rodriguez, Puebla, Mexico - Implicaciones cronoestratigrafias. *Revista Mexicana de Ciencias Geologicas* 16: 187–207.
- McDonald, J. A. 1989. Neotypification of *Ipomoea jalapa* (Convolvulaceae). *Taxon* 38: 135–138.
- McDonald, J. A. 1991. Origin and Diversity of Mexican Convolvulaceae. *Anales del Instituto de Biologia Universidad Nacional Autonoma de Mexico Serie Botanica* 62: 65–82.
- McDonald, J. A. 1994. Convolvulaceae II. In V. Sosa [ed.], *Flora de Veracruz*, vol. 77.
- McDonald, J. A., and D. F. Austin. 1990. Changes and Additions in *Ipomoea* section *Batatas* (Convolvulaceae). *Brittonia* 42: 116–120.
- McDonald, J. A., D. R. Hansen, J. R. McDill, and B. B. Simpson. 2011. A phylogenetic assessment of breeding systems and floral morphology of North American *Ipomoea* (Convolvulaceae). *Journal of the Botanical Research Institute of Texas* 5: 159–177.
- McDonald, J. A., and T. J. Mabry. 1992. Phylogenetic systematics of New World Convolvulaceae based on chloroplast DNA restriction site variation. *Plant Systematics and Evolution* 180: 243–259.
- McNeal, J. R., J. V. Kuehl, J. L. Boore, and C. W. DePamphilis. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* 7: 57.
- Miller, R. E., T. R. Buckley, and P. S. Manos. 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Systematic Biology* 51: 740–753.
- Miller, R. E., J. A. McDonald, and P. S. Manos. 2004. Systematics of *Ipomoea* subgenus *Quamoclit* based on ITS sequence data and a Bayesian phylogenetic analysis. *American Journal of Botany* 91: 1208–1218.

- Miller, R. E., M. D. Rausher, and P. S. Manos. 1999. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and waxy sequences. *Systematic Botany* 24: 209–227.
- Montenegro, Á., C. Avis, and A. Weaver. 2008. Modeling the prehistoric arrival of the sweet potato in Polynesia. *Journal of Archaeological Science* 35: 355–367.
- Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences* 107: 4623–8.
- Muller, J. 1981. Fossil pollen records of extant angiosperms. *Botanical Review* 47: 1–142.
- Noda, N., M. Ono, K. Miyahara, and T. Kawasaki. 1987. Resin glycosides I. Isolation and structure elucidation of orizabin-I, II, III, and IV. Genuine resin glycosides from the root of *Ipomoea orizabensis*. *Tetrahedron* 43: 3889–3902.
- Olmstead, R.G., L. Bohs, H.A. Migid, E. Santiago-Valentin, V.F. Garcia, and S.M. Collier. 2008. A molecular phylogeny of the Solanaceae. *Taxon* 57: 1159–1181.
- Van Ooststroom, S. J., and Leyden. 1953. Convolvulaceae. In C. Van Stennis [ed.], *Flora Malesiana*, 459–489.
- Ozias-Akins, P., and R. L. Jarret. 1994. Nuclear DNA content and ploidy levels in the genus *Ipomoea*. *Journal of the Society for Horticultural Science* 119: 110–115.
- Pares Regali, M. D S., N. Uesugui, and A. D S. Santos. 1974a. Palinologia dos sedimentos meso-cenozoicas do Brasil (I). *Boletim Tecnico da Petrobras, Rio de Janeiro* 17: 177–191.
- Pares Regali, M. D S., N. Uesugui, and A. D S. Santos. 1974b. Palinologia dos sedimentos meso-cenozoicas do Brasil (II). *Boletim Tecnico da Petrobras, Rio de Janeiro* 17: 263–301.
- Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4. Computer program and documentation distributed by the author, website: <http://beast.bio.ed.ac.uk/Tracer> [accessed 22 March 2013].
- Ratan, A. 2009. Assembly algorithms for next-generation sequence data. Ph.D. Dissertation, Pennsylvania State University, University Park, Pennsylvania, USA.
- Rausher, M. D. 2008. Evolutionary transitions in floral color. *International Journal of Plant Sciences* 169: 7–21.
- Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.
- Richards, C. L., B. C. Carstens, and L. Knowles. 2007. Distribution modeling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* 34: 1833–1845.

- Roberty, G. 1952. Genera Convolvulacearum. *Candollea* 14: 11–65.
- Robinson, J. T., H. Thorvaldssdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. Integrative Genomics Viewer. *Nature Biotechnology* 29: 24–26.
- Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–42.
- Roullier, C., L. Benoit, D. B. Mckey, and V. Lebot. 2013. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences* 110: 2205–2210.
- Roullier, C., R. Kambouo, J. Paofa, D. Mckey, and V. Lebot. 2013. On the origin of sweet potato (*Ipomoea batatas* (L.) Lam.) Genetic diversity in New Guinea, a secondary centre of diversity. *Heredity* 110: 594–604.
- Roullier, C., G. Rossel, D. Tay, D. Mckey, and V. Lebot. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Molecular Ecology* 20: 3963–77.
- Schardl, C. L., and K. Clay. 1997. Evolution of mutualistic endophytes from plant pathogens. In G. C. Carroll, and P. Tudzynski [eds.], *The Mycota V: plant relationships*, 221–238. Springer-Verlag, Berlin.
- Schardl, C. L., A. Leuchtman, and M.J. Spiering. 2004. Symbioses of grasses with seedborne fungal endophytes. *Annual Review of Plant Biology* 55: 315–340.
- Schardl, C. L., C. A. Young, U. Hesse, S. G. Amyotte, K. Andreeva, P. J. Calie, D. J. Fleetwood, et al. 2013. Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *Plos Genetics* 9: e1003323.
- Shinners, L. 1979. Convolvulaceae. In D. S. Correll, and M. C. Johnston [eds.], *Manual of the Vascular Plants of Texas*. The University of Texas at Dallas, Texas Research Foundation, Richardson, Texas.
- Smith, S.D., and D.A. Baum. 2007. Systematics of *Iochrominae* (Solanaceae): Patterns in floral diversity and interspecific crossability. *Acta Horticulturae* 745: 241–254.
- Smith, S. D., R. E. Miller, S. P. Otto, R. G. FitzJohn, and M. D. Rausher. 2010. The effects of flower color transitions on diversification rates in morning glories (*Ipomoea* subg. *Quamoclit*, Convolvulaceae). *Proceedings of the Darwin 2000 Beijing International Conference* 186–210.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–90.

- Stefanovic, S., D. F. Austin, and R. G. Olmstead. 2003. Classification of Convolvulaceae: A phylogenetic approach. *Systematic Botany* 28: 791–806.
- Stefanovic, S., L. Krueger, and R. G. Olmstead. 2002. Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *American Journal of Botany* 89: 1510–1522.
- Steiner, U., S. Leibner, C. L. Schardl, A. Leuchtman, and E. Leistner. 2011. *Periglandula*, a new fungal genus within the Clavicipitaceae and its association with Convolvulaceae. *Mycologia* 103: 1133–1145.
- Streisfeld, M.A., and J.R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59: 2548–2559.
- Streisfeld, M. A., and M. D. Rausher. 2009. Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytologist* 183: 751–763.
- Streisfeld, M. A., and M. D. Rausher. 2010. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* 65: 629–642.
- Swofford, D. L. 2003. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thorvaldsson, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14: 178-192.
- Tripp, E. A., and P. S. Manos. 2008. Is Floral Specialization an Evolutionary Dead-End? Pollination System Transitions in *Ruellia* (Acanthaceae). *Evolution* 62: 1712-1737.
- Verdcourt, B. 1963. Convolvulaceae. In C. E. Hubbard, and E. Milne-Redhead [eds.], *Flora of Tropical East Africa*. Whitefriars Press, London.
- Verdcourt, B. 1957. Typification of the Subdivisions of *Ipomoea* L. (Convolvulaceae) with particular regard to the East African species. *Taxon* 6: 150–152.
- Wessinger, C. A., and M. D. Rausher. 2012. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany* 63: 5741–5749.
- Wikström, N., V. Savolainen, and M. W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society* 268: 2211–2220.
- Wilkin, P. 1999. A morphological cladistic analysis of the Ipomoeae (Convolvulaceae). *Kew Bulletin* 54: 853–876.
- Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Zufall, R.A., and M.D. Rausher. 2004. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* 428: 847–850.

Table 2.1 – Mean, minimum and maximum node ages of clades denoted in figure 2.2. Minimum and maximum ages represent 95% highest posterior densities.

<b>Clade</b>	<b>Mean Age</b>	<b>Min Age</b>	<b>Max Age</b>
Murucoides	4.82	1.08	9.62
Batatas	12.43	6.44	19.34
Pes-caprae	13.24	5.19	22.05
Quamoclit	21.28	12.05	31.04
Cairica	22.21	12.39	32.59
Astripomoeinae	23.39	13.51	34.44
Obscura	22.75	12.47	34.32
Pes-tigridis	22.79	12.27	34.49
Argyreinae	26.38	15.01	38.46
Ipomoeae	34.97	21.08	49.64
Merremieae	49.34	41.20	61.99
Convolvulaceae	55.29	47.80	69.98
Solanaceae	44.71	23.00	72.82

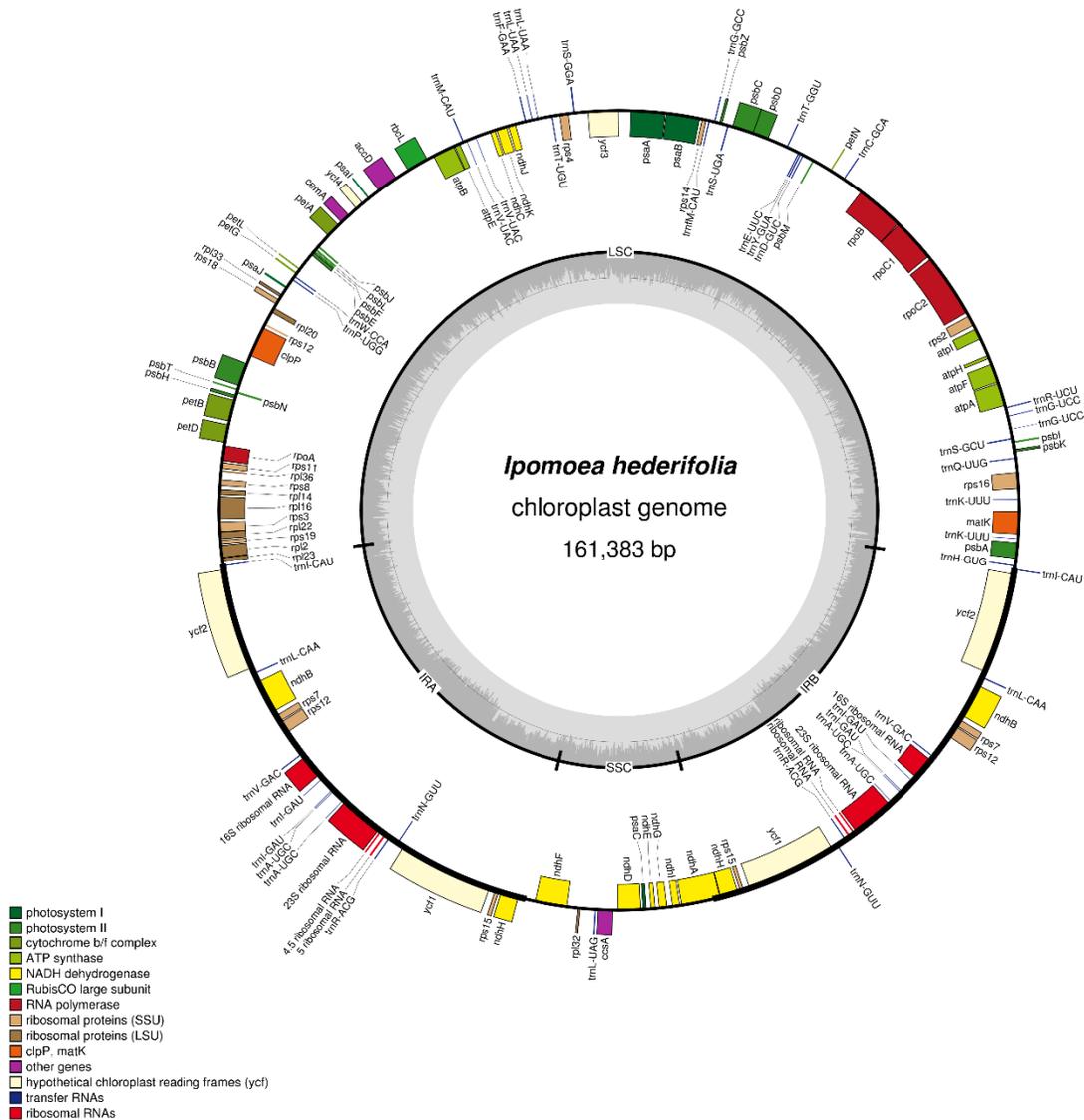


Figure 2.1 – Chloroplast genome of *Ipomoea hederifolia*. The outer circle shows positions of genes and the large single copy (LSC), small single copy (SSC), and two inverted repeat (IR<sub>A</sub> and IR<sub>B</sub>) regions. The inner circle is a graph depicting GC content across the genome (dark grey bars = percent GC). Plastome maps were generated in OGDraw v1.2 (Lohse et al. 2007, 2013).

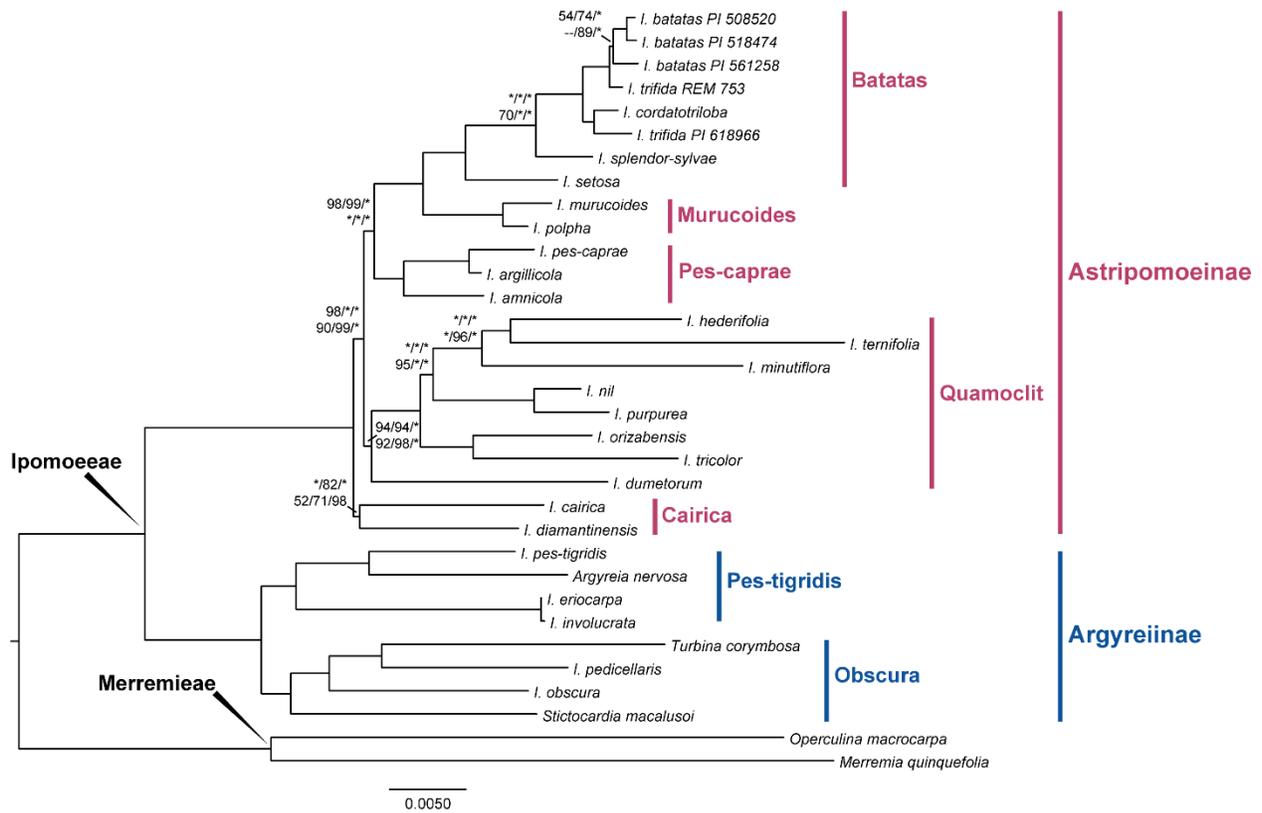


Figure 2.2 – Phylogeny of the Ipomoeae based on whole chloroplast genome sequences. The second inverted repeat region was removed for analyses. The topology shown is from a maximum likelihood analysis in RAxML of the mauve alignment. Numbers behind nodes are maximum parsimony bootstrap (MP), maximum likelihood bootstrap (ML), and Bayesian posterior probability (PP) values for the mauve and SATé alignments. Nodes without numbers or with an asterisk (\*) received 100% bootstrap and PP support in all analyses. Top numbers are Mauve MP, ML, and PP values. Lower numbers are SATé MP, ML, and PP values. Pink bars to the right are well-supported lineages within the Astripomoeinae; blue bars are lineages within the Argyreiinae.

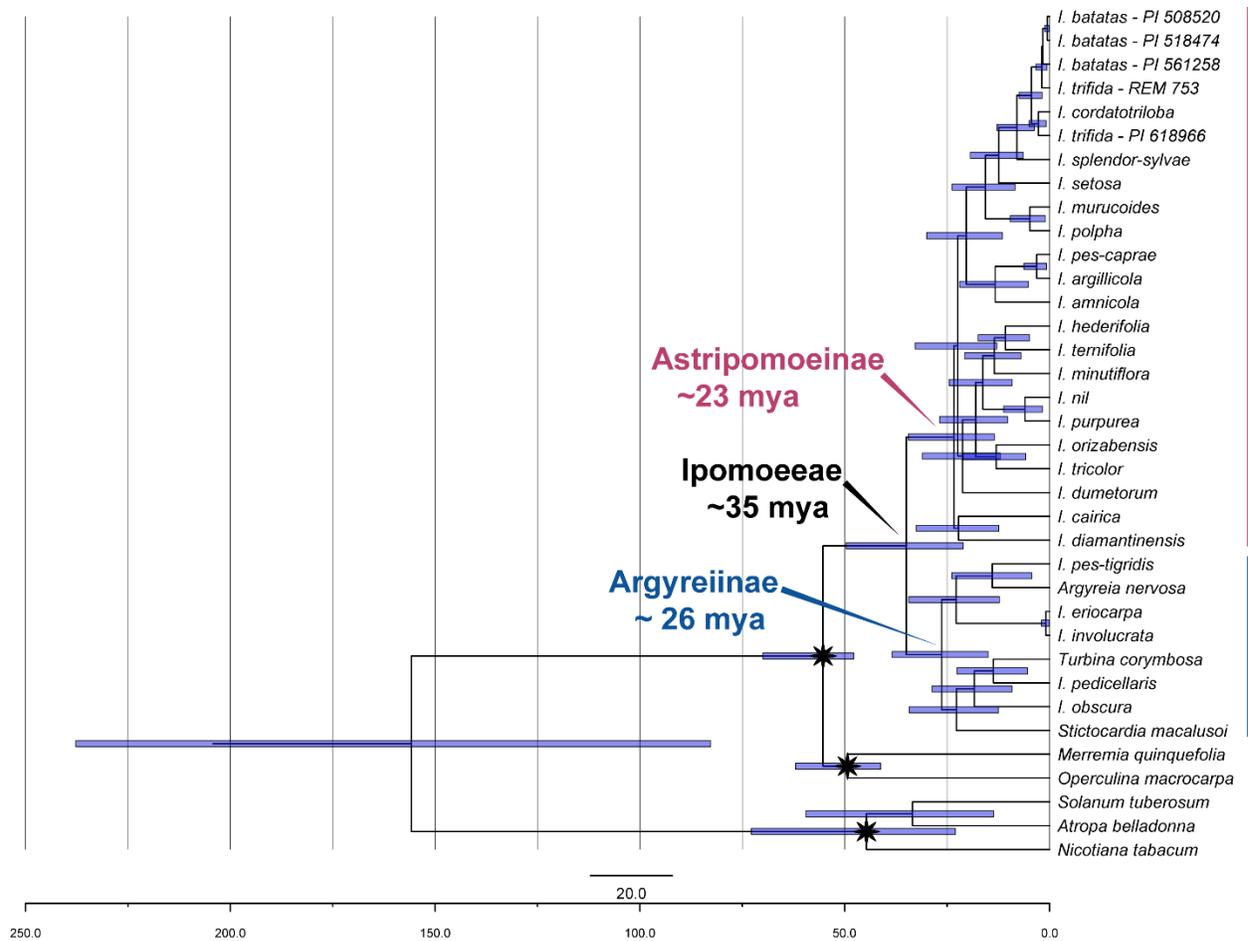


Figure 2.3 – Results of the divergence time analysis. Blue bars around nodes are 95% highest posterior densities. Nodes are placed based on the mean node age. Stars denote nodes calibrated with fossil pollen.

CHAPTER III:

THE EVOLUTIONARY HISTORY OF THE SWEETPOTATO COMPLEX, *IPOMOEA* SERIES *BATATAS* (CHOISY)

D. F. AUSTIN, USING TARGET ENRICHMENT<sup>2</sup>

<sup>2</sup>Eserman, Lauren A., Jim H. Leebens-Mack. To be submitted to American Journal of Botany.

## **Abstract**

Sweetpotato (*Ipomoea batatas* (L.) Lam.) is one of the most important crop species worldwide for human nutrition. The large storage roots provide a critical source of carbohydrates and vitamin A, especially in developing countries. Sweetpotato production is currently limited by a small number of improved accessions. However, the wild relatives of crop species have the potential to be stores of agronomically important traits. The relationships between sweetpotato and its wild relatives, the Batatas complex, is currently poorly understood. Most taxa examined were diploid with the exception of cultivated sweetpotato (6x) and *Ipomoea tabascanana* (4x). Phylogenomic analyses recovered four major lineages in the Batatas complex. Sweetpotato was closely allied with *I. trifida*, and *I. tabascanana* was found to be closely related to *I. triloba*. Hybridization analysis suggests that cultivated sweetpotato has hybrid ancestry, with parentage from *I. ramosissima* and either *I. triloba* or *I. cordatotriloba*. Two tests for introgression reveal a single ancient hybridization event in the ancestor of the primarily North American and Mexican clade. Phylogenetic results presented here advance understanding the relationships among sweetpotato and its wild relatives. Furthermore, these results suggest there were at least two independent origins of polyploidy in the Batatas complex. Ancient hybridization and polyploidization certainly played an important role in the evolutionary history of the Batatas complex. These results will advance sweetpotato breeding efforts.

## **Introduction**

The sweetpotato (*Ipomoea batatas* (L.) Lam.) complex, or the Batatas complex, is a diverse group of species comprising sweetpotato and its wild relatives which are distributed throughout tropical, subtropical and temperate regions of North and South America (Austin 1978; Austin and Huáman 1996; Khoury et al. 2015). The Batatas complex currently consists of

14 named species which were circumscribed based on vegetative and floral morphology (Austin 1978, 1988a; McDonald and Austin 1990) and unnamed but morphologically distinct populations (Duncan and Rausher 2013, *pers. obs.*). Sweetpotato (*Ipomoea batatas*) is one of the earliest domesticated plant species (Ugent and Peterson 1988) and remains one of the most important crops worldwide for human nutrition because of the large quantities of vitamin A and carbohydrates produced in the storage roots (Lebot 2009; FAO 2016). The vast diversity in storage root shapes, colors and uses in sweetpotato make it a popular vegetable (Loebenstein 2009). Ornamental sweetpotato vines are also commonly used in planters or for ground cover (Winslow 2012). Despite the economic importance of sweetpotato, the relationships among members of the Batatas complex is poorly understood. In this study, we examine the utility of target enrichment to resolve relationships and test for both ancient hybridization and ongoing gene flow among the wild relatives of sweetpotato.

Numerous attempts have been made to describe the evolutionary history of the sweetpotato complex; however, this has proven difficult due to ploidy differences among individuals and putative homoploid and polyploid hybrid species. Austin (1988) hypothesized evolutionary relationships and hybridization events in the complex using morphological similarity among taxa. Two attempts have been made using molecular data to resolve relationships. Jarret et al. (1992) used RFLP data to generate a tree of the Batatas complex, and Rajapakse et al. (2004) attempted to reconstruct the phylogeny of the sweetpotato complex using beta-amylase sequences. The phylogenetic inferences from each of these were highly incongruent, with the largest disagreements being between the trees estimated from morphological and molecular data.

No study to date has explicitly tested for hybridization or reticulate evolution in the Batatas complex. In many groups of species, the true evolutionary relationships among taxa may be best represented by a network depicting recent or ancient hybridization events among lineages (Huson and Bryant 2006; Yu et al. 2014). These processes may have been important in the Batatas complex, given the fact that many species are able to hybridize with one another (e.g. Abel and Austin 1981; Oracion et al. 1990; Diaz et al. 1996; Cao et al. 2009). Some species, such as *Ipomoea leucantha* and *I. grandifolia* are also thought to be the result of homoploid hybrid speciation, and sweetpotato is thought to have arisen through either allopolyploidization (Abel and Austin 1981; Austin 1988b) or autopolyploidization (Kriegner et al. 2003; Cervantes-Flores et al. 2007). In addition, there is some evidence from microsatellites of ongoing gene flow between sympatric *I. cordatotriloba* and *I. lacunosa* (Duncan and Rausher 2013). The observation that many species in the Batatas complex appear morphologically similar and show evidence of hybridization may be the result of ongoing or ancient (i.e. no longer ongoing) hybridization. Distinguishing between ongoing versus ancient hybridization is difficult, but coalescence-based methods are available to test these hypotheses (Than et al. 2008; Yu et al. 2014; Kubatko and Chifman 2015; Yu and Nakhleh 2015). Here we employ these methods to decipher complicated genetic patterns and advance understanding of the complex evolutionary history of the Batatas complex.

The goals of this study are: (1) to test the utility of target enrichment in the sweetpotato complex, (2) to infer incomplete lineage sorting within the species phylogeny of the sweetpotato complex, (3) to examine the effect of hybridization or reticulate evolution in this group while also accounting for incomplete lineage sorting, and (4) to estimate the phylogenetic placement of polyploid taxa in the Batatas complex. Deep coalescence, or incomplete lineage sorting (ILS), is

a known source of gene tree discordance, especially in very young rapidly diversifying groups (Pamilo and Nei 1988; Maddison 1997; Edwards 2009). Sequencing and analysis of DNA libraries enriched for single copy genes has been a powerful approach for resolving species relationships among closely related species in the face of ILS (Heyduk et al. 2015; Stephens et al. 2015a, 2015b). The Batatas complex is a relatively young clade which diversified c. 12 mya (Eserman et al. 2014), so ILS is likely to have resulted in extensive gene tree/species tree discordance. In addition, as discussed above, we expect to find some level of gene tree discordance due to either recent or ancient hybridization. We are utilizing two methods to assess the effect of hybridization while also accounting for incomplete lineage sorting. PhyloNet identifies reticulation events given a set of gene trees (Than et al. 2008; Yu et al. 2014; Yu and Nakhleh 2015), and HyDe identifies sets of putative hybrid taxa given a multi-locus gene alignment (Kubatko and Chifman 2015). If the hybridization events were recent (or there is ongoing gene flow among populations), we would expect PhyloNet to identify multiple hybridization events near the tips of the tree. We would also expect the HyDe results to show that the parents of the inferred hybrid taxa were from nearby geographic areas. In contrast, if hybridization events were ancient, we would expect PhyloNet to implicate one or a few hybridization events deep in the tree, and HyDe would likely identify multiple hybrid tips as the products of a single shared, ancient hybridization event. Further, the hybrid parents implicated by HyDe would not necessarily be from the same geographic region. Finally, we expect that cultivated *I. batatas* will show a signature of hybridization if it is of allopolyploid origin, with some contribution from *I. trifida* or a close relative. We predict that cultivated *I. batatas* will be phylogenetically close to *I. trifida* as in previous studies (Austin 1988b; Jarret et al. 1992;

Rajapakse et al. 2004), and we also expect to find some contribution of *I. triloba* in the hybrid evolutionary history of *I. batatas*, as predicted by Austin (1988b).

## **Methods**

### *Taxon sampling*

Fourteen accessions were selected to represent the breadth of geographic and morphological diversity in the Batatas complex (Table 1). Taxon sampling was also guided by a preliminary phylogeny of the Batatas complex based on three nuclear intron sequences and one chloroplast intergenic spacer (Tiley et al. *in prep*). In addition, two accessions of *Ipomoea triloba* and two accessions of *I. trifida* were included in our analyses because previous studies have shown that these species may not be monophyletic, i.e. samples with similar morphology may represent distinct evolutionary lineages (Tiley et al. *in prep*; Eserman et al. 2014). Seeds and cuttings were obtained from USDA GRIN and from the research collections of Drs. Rick Miller and Michael T. Clegg. Seeds of the outgroup species *I. setosa* were obtained from B&T World Seeds.

### *RNA bait design*

Targeted sequence capture requires a set of RNA baits complimentary to exon sequences in the species of interest. To generate this bait set, we used gene sequences from the two published *Ipomoea trifida* genomes (Hirakawa et al. 2015) and an unpublished *I. triloba* genome made available to us by the International Sweetpotato Genome Initiative (<http://sweetpotato.plantbiology.msu.edu>). Exon sequences from all three genomes were then sorted into orthologous groups, including a set of known to be single copy in eudicots (Duarte et al. 2010; Amborella Genome Project 2013). We then identified a set of genes which were verified as being single copy in three morning glory genome assemblies (Hirakawa et al. 2015;

<http://sweetpotato.plantbiology.msu.edu>). Intron-exon boundaries were determined in gene models from the more complete of the two *I. trifida* genome assemblies, Mx23Hm (Hirakawa et al. 2015). RNA baits were designed to tile across exon sequences in the Mx23Hm genome with 60 bp overlap between 120 bp biotinylated oligonucleotides (Mycroarray, Ann Arbor, MI). In total, we targeted 1953 exons distributed among 490 genes.

#### *Genome size measurements*

Fresh leaf tissue was sent to the Flow Cytometry Lab at the Benaroya Research Institute at Virginia Mason. Intact nuclei were isolated and subjected to flow cytometric analysis for nuclear DNA content. Each sample was measured in triplicate. Mean nuclear DNA content was measured in pg/2C. Chicken red blood cell nuclei (2C=2.5 pg/2C) were used as an internal standard, the same standard used in a previous flow cytometric analysis of genome size in these species (Ozias-Akins and Jarret 1994). Ploidy and nuclear DNA content have been found to be highly correlated (Ozias-Akins and Jarret 1994); therefore, ploidy level was estimated from flow cytometric measurements of nuclear DNA content.

#### *Library preparation and sequencing*

Total genomic DNA was isolated from fresh or dried leaf tissue using a modified CTAB protocol (Doyle 1987; Storchova et al. 2000). DNA samples were sheared with a Covaris sonicator to an average insert size of 500 bp. Libraries were prepared using either the KAPA HTP DNA Library Kit for Illumina (Kapa Biosystems, Wilmington, MA) or using an in-house protocol modified from (Fisher et al. 2011). Library concentration was determined using quantitative real-time PCR. The length distribution of libraries was determined on the Agilent Bioanalyzer 2100 using the DNA 1000 kit (Agilent Technologies, Santa Clara, CA). Four barcoded DNA libraries were pooled in equal concentration and used as input for the

hybridization reaction, following guidelines in the MyBaits protocol (version 3). RNA baits were hybridized to pooled DNA libraries, and the biotinylated baits bound to DNA were captured with streptavidin beads (ThermoFisher Scientific, Waltham, MA). Concentration of enriched libraries was determined by quantitative real-time PCR and Qubit dsDNA high-sensitivity assay kit (ThermoFisher Scientific, Waltham, MA). Enriched libraries were pooled to a final concentration of 10 nM and sequenced on the Illumina NextSeq platform with 150 bp paired-end reads.

### *Sequence assembly*

Sequence assembly generally followed the reads2trees pipeline (Heyduk et al. 2015). Reads were sorted by barcode. Trimmomatic (Bolger et al. 2014) was used to trim adapter and barcode sequences as well as low quality ends from reads. Reads which were less than 40 bp in length after adapter and barcode trimming were removed in Trimmomatic. Trimmed reads were assembled *de novo* in Trinity version 2.0.6 (Haas et al. 2013). Contigs assembled in Trinity were then matched to the exons used to design baits using BLAST. Any time more than one Trinity contig matched an exon in the reference set, these contigs were removed from the analysis. These contigs can represent paralogs or alleles of a gene, but these could not be separated in this assembly. Therefore, these contigs were removed to only retain single-copy genes. However, to verify that we were obtaining single-copy genes in the two polyploid taxa, we used Bowtie2 (Langmead and Salzberg 2012) to map cleaned, filtered reads to the single copy gene assemblies of *Ipomoea batatas* cultivar (cv.) Tinian and *I. tabascanana*. Read mapping was inspected in Geneious (Kearse et al. 2012). Any gene exhibiting evidence for more than two alleles was removed from further analysis. Finally, sequences belonging to the single copy gene families were extracted from two published *Ipomoea trifida* genomes (Hirakawa et al. 2015) and one *I. trifida* and one *I. triloba* genome from the International Sweetpotato Genome Initiative

(<http://sweetpotato.plantbiology.msu.edu>) using BLAST. Multiple sequence alignments for each single copy gene family were aligned using PRANK (Loytynoja and Goldman 2005; Löytynoja and Goldman 2008), and alignments were filtered using Gblocks (Castresana 2000; Talavera and Castresana 2007) to remove poorly aligned regions from the alignment. We then used the resulting gene family alignments to make five datasets for phylogenetic analyses: (1) one with only diploid taxa, (2) one with diploid taxa and *I. batatas* cv. Tinian filtered by BLAST, (3) one with diploid taxa and *I. batatas* cv. Tinian filtered to remove loci exhibiting evidence in read mapping analysis for possible collapsing of paralogous genes (see above), (4) one with diploids and *I. tabascana* filtered by BLAST, and (5) one with diploids and *I. tabascana* filtered by read mapping.

#### *Phylogenetic analysis*

Gene trees were estimated separately for each gene alignment in RAxML, and bootstrap support was calculated from 100 bootstrap replicates using the rapid bootstrapping algorithm (Stamatakis 2014). A concatenated alignment of all genes was also used to estimate species relationships using RAxML. *Ipomoea setosa* was used as the outgroup in all phylogenetic analyses. Trees were also generated in ASTRAL-II version 4.10.12 (Mirarab and Warnow 2015) and SVDQuartets (Chifman and Kubatko 2014) to account for possible incomplete lineage sorting. These programs were chosen because they account for ILS in different ways. ASTRAL-II optimizes quartet frequencies at each node across all estimated gene trees to reconstruct a species tree. However, ASTRAL-II assumes that gene tree estimates are true trees, i.e. estimated without error, and this assumption is almost certainly violated. In contrast, SVDQuartets uses single nucleotide polymorphism (SNP) data to estimate a species tree. One assumption of SVDQuartets is that SNPs are independent of one another. However, this assumption is violated

in our analyses because we are using a concatenated alignment of multiple genes, and each gene contains multiple SNPs. ASTRAL-II and SVDQuartets have both been shown to be robust to violation of their respective simplifying assumptions; however, we must acknowledge that violating these assumptions may mislead our phylogenetic inferences. Therefore, we apply both methods to estimate species trees to account for ILS. Both methods also allow for bootstrap resampling to assess branch support but apply bootstrapping in different ways. ASTRAL-II bootstrap resamples gene trees (Sayyari and Mirarab 2016), while SVDQuartets resamples sites in the concatenated alignment with replacement (Chifman and Kubatko 2014).

### *Inferring hybridization*

Two approaches which account for both incomplete lineage sorting and hybridization were also applied to these data. PhyloNet (Than et al. 2008; Yu and Nakhleh 2015) was used to estimate a phylogenetic network under zero, one, two, and three reticulation scenarios using the InferNetwork\_MPL option. The zero reticulation scenario run in PhyloNet is equivalent to estimating a phylogenetic tree under the multi-species coalescent. Maximum likelihood gene trees estimated using RAxML were used as input in PhyloNet. Five independent analyses of each reticulation scenario were carried out to best traverse the complex parameter space. The phylogenetic network with the best likelihood score, which was significantly different from the likelihood of a zero-reticulation network (i.e. the ASTRAL-II and SVDQuartets trees) was chosen as the most likely network given the gene trees.

HyDe (Kubatko and Chifman 2015) was also used to detect hybridization. HyDe uses phylogenetic invariants (Felsenstein and Cavender 1987) to estimate hybrid taxa and potential parental taxa from the concatenated multi-locus gene alignment. Using this approach, at least 250 sites were required to share a particular site pattern in order for HyDe to test for

hybridization. A p-value of  $6.87 \times 10^{-5}$  was applied to correct for multiple comparisons using the Bonferroni method (Kubatko and Chifman 2015). To further explore the results of the HyDe analysis, we calculated pairwise distance in a subset of hybrids and inferred parental taxa for the diploid dataset using the DNASTatistics module within BioPerl (D\_Uncorrected option). Both HyDe and PhyloNet were run on the diploid dataset as well as the two datasets including hexaploid *I. batatas* or *I. tabascana*. When HyDe was run on the polyploid datasets, the p-value was adjusted to be  $5.50 \times 10^{-5}$  to account for multiple comparisons with 16 taxa. When HyDe was run on the dataset that included hexaploid *I. batatas* and was filtered by read mapping, 150 sites were required to share a site pattern for the test of hybridization to accommodate the loss of ca. 60% of the SNP sites after filtering.

## Results

### *Genome size*

We measured genome size on twelve of the fourteen samples sequenced for this study. The samples ranged from 0.81 ( $\pm 0.025$ ) pg DNA/2C nucleus to 2.69 ( $\pm 0.044$ ) pg DNA/2C nucleus (Table 2). Genome size of *Ipomoea setosa*, the outgroup species, was 1.52 ( $\pm 0.041$ ) pg DNA/2C nucleus (Table 2).

### *Sequencing results*

Results of sequencing and assembly are reported in Table 3. The total number of unfiltered reads varied by library, ranging from 275,692 to 3,227,800 reads per library. Filtering by read length in Trimmomatic resulted in an overall reduction of 59 to 665 reads per library. Furthermore, we were able to assemble between 180 to 382 genes per sample. Mean exon coverage was greater than 50x for most samples. Mean exon coverage ranged from 13.78x to

287.71x. Mean intron coverage was generally lower than exon coverage and ranged from 6.72x to 63.50x coverage.

### *Data matrices*

The concatenated gene matrix for the 15 diploid taxa was 722,619 nucleotides in length and contained 38,668 variable sites which were parsimony uninformative and 16,400 parsimony informative sites (Table 4). Parsimony informative sites are SNP loci where a minor allele was found in two or more samples; sites with a minor allele in only one sample are parsimony uninformative. The data matrix containing tetraploid *Ipomoea tabascanana* which was filtered only by BLAST contained 738,135 nucleotides, 38,904 were variable but parsimony uninformative and 17,437 were parsimony informative. When reads were mapped to the BLAST filtered genes and genes showing more than two haplotypes in the reads, this resulted in a reduction in genes from 366 to 351. The matrix including *I. tabascanana* which had been further filtered by read-mapping contained 711,058 nucleotides, where 37,094 were variable but parsimony uninformative and 16,599 were parsimony informative. The data matrix which contained hexaploid *I. batatas* which had been filtered only by BLAST contained 710,446 nucleotides, where 39,050 were parsimony uninformative and 17,111 were parsimony informative. Removal of genes with greater than two haplotypes resulted in a reduction from 366 to 261 genes. The matrix containing hexaploid *I. batatas* which had been further filtered by read-mapping contained 456,369 nucleotides, where 25,621 were parsimony uninformative and 11,150 were parsimony informative. The three datasets which had been filtered only by BLAST contained the exact same 366 genes.

### *Phylogenetic results*

For the diploid taxa, the concatenation, ASTRAL-II and SVDQuartets trees show identical topologies (Figure 3.1). The Batatas complex was recovered as monophyletic in all analyses. Further, three major clades were found within the Batatas complex, denoted by the red, yellow and blue bars on Figure 3.1. *Ipomoea lacunosa* was recovered as part of the red clade in all three trees with low bootstrap support and on a very short branch; therefore, it was considered to be representative of a separate lineage, denoted by the green bar (Figure 3.1). Bootstrap support on the branch leading to *I. lacunosa* and the red clade was the lowest and was 67% in the concatenation tree, 84% in the ASTRAL-II tree, and 68% in the SVDQuartets tree. The branch leading to the last common ancestor of the red, green, yellow and blue clades showed 100% bootstrap support in the concatenation and ASTRAL-II trees but had bootstrap support of 96% in the SVDQuartets tree. The normalized quartet score, an indicator of gene tree discordance, for the ASTRAL-II tree was 0.6558 for the entire tree, and the quartet score for individual clades ranged from 0.3529 on the branch leading to the red clade and *I. lacunosa* to 0.7280 at the base of the red, yellow, and green lineages. These relatively low normalized quartet scores implicate extensive ILS due to rapid diversification.

The ASTRAL-II trees for the dataset including *I. tabascanana* and genes filtered by BLAST (i.e. genes from an individual with more than one BLAST hit to the reference gene set) and by read mapping resulted in trees with identical topologies (Figure 3.2). In both trees, *I. tabascanana* was recovered as sister to *I. triloba* NSP323. The ASTRAL-II trees for the dataset containing *I. batatas* cv. Tinian including genes filtered by BLAST and read-mapping showed differing topologies. One difference is the placement of *I. lacunosa*, which was sister to the red clade in the tree containing genes filtered by BLAST and the trees including only diploid taxa. However,

*I. lacunosa* was sister to the yellow clade in the tree generated with the dataset filtered by read mapping with low bootstrap support (69%, Figure 3.2). In addition, the placement of *I. batatas* cv. Tinian differed in the two trees (Figure 3.2). *Ipomoea batatas* cv. Tinian was recovered as most closely related to *I. ramosissima* and this clade was sister to an *I. trifida* accession from Costa Rica in the dataset where genes were filtered by BLAST. However, in the dataset where genes were filtered by read mapping, *I. ramosissima* and *I. trifida* from Costa Rica were sister taxa, with very low bootstrap support, and *I. batatas* cv. Tinian was sister to this clade (Figure 3.2).

#### *Inference of hybridization*

HyDe (Kubatko and Chifman 2015) was used to identify potential hybrid and parental taxa in the diploid dataset. The HyDe analysis using only diploid taxa inferred six hybrid taxa in three clades: 2 of 3 samples in the red clade, 3 of 3 samples in the yellow clade, and 1 of 5 samples in the blue clade (Figure 3.3). The inferred parents of the hybrid taxa showed no obvious pattern and were distributed across the phylogeny, but in no case were both inferred parents identified within the same clade (Figure 3.3). PhyloNet (Than et al. 2008; Yu and Nakhleh 2015) was also used to identify patterns of reticulate evolution in the Batatas complex. The network with the highest likelihood score showed a single reticulation event in the ancestor of the clade containing the red, yellow and green lineages (Figure 3.4). The inferred reticulation was estimated as a hybrid of the ancestor of the blue clade and the ancestor of the red, green, and yellow clade. Finally, we examined pairwise genetic distance in a subset of taxa identified as hybrids and their inferred parents. These results illustrate that the inferred parent residing in the same clade on the phylogeny as the putative parent showed a high degree of similarity across all

loci, and the inferred parent from a different clade exhibited a higher degree of divergence between most orthologous genes (Figure 3.5).

Additionally, HyDe and PhyloNet were run on the BLAST filtered and read mapping filtered polyploid datasets. When *I. batatas* cv. Tinian was included in the dataset, the HyDe analysis of the BLAST filtered genes inferred 6 hybrid taxa, and the analysis of read mapped filtered genes inferred 4 hybrid taxa (Figure 3.6). When *I. tabascana* was included in the dataset, the HyDe analysis of the BLAST filtered genes inferred 8 hybrid taxa, and the genes filtered by read mapping inferred 9 hybrid taxa (Figure 3.6). In all cases, inferred parents of the hybrid taxa were from very different parts of the tree and often were from different geographic locations. The PhyloNet results for the dataset including *I. batatas* cv. Tinian resulted in a maximum likelihood network with two reticulations, one in the ancestor of the red, yellow and green lineages, and the other in the ancestor of cultivated sweetpotato (Figure 3.7). When *I. tabascana* was included in the dataset, the maximum likelihood network inferred three reticulation events, all concentrated in the blue clade (Figure 3.7).

## **Discussion**

### *Inference of ploidy*

A prior study found a linear relationship between ploidy and nuclear DNA content measured by flow cytometry (Ozias-Akins and Jarret 1994), which compared ploidy inferred from chromosome counts measured by root tip squashes to ploidy inferred from flow cytometry. They found that tetraploids had approximately two times the DNA content of diploids, and hexaploids had approximately 1.25-1.5 times the DNA content of tetraploids (Ozias-Akins and Jarret 1994). In the samples measured in this study, *Ipomoea tabascana*, which was previously characterized as tetraploid, has approximately two times the DNA content of most samples.

Therefore, we infer that *Ipomoea tabascanana* is tetraploid, and the samples ranging from 0.81-1.06 pg DNA/2C nucleus are diploid. Furthermore, cultivated sweetpotato (cultivar Tinian) contains 1.37 times the DNA content compared to tetraploid *I. tabascanana*. Therefore, we can confirm in this study that sweetpotato cultivar Tinian is hexaploid. Genome size measured by flow cytometry has never been reported for the outgroup species, *I. setosa*. In this study, we found that the *I. setosa* genome size is 1.52 ( $\pm 0.041$ ) pg DNA/2C nucleus. Previous work has shown that *I. setosa* is diploid with 30 chromosomes (Wolcott 1937; Sharma and Datta 1958). The higher diploid genome size in *I. setosa* may be due to factors such as LTR-retrotransposon proliferation (Bennetzen 2002). Overall, genome size measurements were found to be lower than in previous studies of these species (Arumuganathan and Earle 1991; Ozias-Akins and Jarret 1994); however, the values were internally consistent in this study (Table 2).

#### *Target enrichment*

There have been three previous attempts to reconstruct the phylogeny of the Batatas complex. The first used only morphological characters (Austin 1988b), a second study applied RFLP data (Jarret et al. 1992), and the third used intron and exon sequences of beta-amylase (Rajapakse et al. 2004). When compared with largest dataset in the latter study, the present research includes more than 650 times the DNA sequence data and 40 times the number of parsimony informative sites, illustrating the power of target enrichment to generate a large number of single-copy, orthologous genes for phylogenetic reconstruction.

#### *Phylogenetic relationships in the Batatas complex*

When we compare the topologies to previous studies, Austin (1988) found that *I. trifida* and *I. ramosissima* belonged to a clade, and *I. cordatotriloba* and *I. tenuissima* belonged to a clade. However, Austin (1988) hypothesized that *I. cordatotriloba* and *I. lacunosa* are sister taxa,

but the present results are inconsistent with this topology. Consistent with this study, Jarret et al. (1992) and Rajapakse et al. (2004) recovered *I. batatas* and *I. trifida* as members of the same clade. Jarret et al. (1992) further found *I. cordatotriloba* and *I. tenuissima* belong to the same clade. Further, the relationships among *I. cordatotriloba*, *I. tenuissima*, *I. lacunosa*, and *I. triloba* recovered in Jarret et al. (1992) match the present topology. However, Jarret et al. (1992) and Rajapakse et al. (2004) recovered *I. tabascana* as most closely related to *I. batatas*; whereas, we recovered *I. tabascana* as closely related to *I. triloba*. Finally, the placement of *I. ramosissima* differed between the present study and both Jarret et al. (1992) and Rajapakse et al. (2004). We found *I. ramosissima* to be nested within the *I. trifida* accessions and in one instance the sister taxon to cultivated *I. batatas*. However, Jarret et al. (1992) recovered *I. ramosissima* as sister to the rest of the Batatas complex, and Rajapakse et al. (2004) found *I. ramosissima* was sister to *I. splendor-sylvae* (= *I. umbraticola*), which was sister to the rest of the Batatas complex.

It must be noted that morphological characters used to diagnose species in the Batatas complex are highly variable, and individuals exhibiting intermediate phenotypes are often found (Austin 1978, 1988b; *pers. obs.*; *pers. comm.*, Dr. Rick Miller). For example, sepal shape, flower size, and the number of raised veins on sepals are often diagnostic of species (Austin 1978, 1988a; McDonald and Austin 1990). These characters, especially flower size, often exhibit high phenotypic plasticity, such that plants collected in the field differ considerably from plants grown in the greenhouse (*pers. comm.*, Dr. Rick Miller). Taken together, much denser taxon sampling, careful documentation of characters which do not exhibit environmental plasticity, taxonomic reassessment, and a deeper understanding of how hybridization has shaped the evolutionary history of the Batatas complex will be crucial to understanding relationships among these species.

### *Reticulate evolution*

Recently radiated species groups present unique challenges in phylogenomics. Gene tree discordance is pervasive and well-documented in phylogenomic studies (e.g. (Refulio-Rodriguez and Olmstead 2014; Heyduk et al. 2015; Stephens et al. 2015a; Comer et al. 2016). To account for the discordance of gene trees, many phylogenetic studies employ methods that account for incomplete lineage sorting (ILS), a phenomenon where ancestral allelic diversity is retained between speciation events, and gene trees reflect mutational histories that predate the speciation events being inferred. As a consequence, ILS results in gene phylogenies that are not necessarily concordant with the history of population divergence and speciation (Degnan and Rosenberg 2009; Edwards 2009). The signature of ILS is especially strong when branch lengths are short or effective population sizes are especially large (Degnan and Rosenberg 2009). A wealth of methods have been developed that model the multispecies coalescent which can account for ILS in multilocus gene datasets (e.g. (Liu et al. 2010; Bouckaert et al. 2014; Chifman and Kubatko 2014; Mirarab and Warnow 2015).

Accounting for ILS alone may be insufficient to represent the true evolutionary history of a group of species where the true species relationships are best described by a network with recent and/or ancient reticulation events (Huson and Bryant 2006; Yu et al. 2014; Wen et al. 2016). Methods that account for both ILS and hybridization or reticulation have recently been developed, but these methods are often computationally intensive in the case of PhyloNet (Than et al. 2008; Yu et al. 2014; Yu and Nakhleh 2015) or are limited in the inferences they can draw in the case of HyDe (Kubatko and Chifman 2015). PhyloNet has the power to detect recent and/or ancient reticulation events by estimating the maximum likelihood phylogenetic network given a set of gene trees, but this method is computationally intensive, even when applying the

speedier maximum pseudo-likelihood approach. In contrast, HyDe is computationally very fast but does not estimate a network. Instead, HyDe infers putative hybrid individuals and the parental taxa given a multilocus gene alignment or a set of unlinked SNPs. Therefore, we applied both HyDe and the maximum pseudolikelihood approach in PhyloNet to test for either recent or ancient reticulation in the evolutionary history of the Batatas complex.

Results of both the HyDe and PhyloNet analyses suggest that ancient reticulation occurred in the ancestor of the red, yellow, and green lineages and rapid diversification occurred after this reticulation event (Figure 3.3, 3.4). The network from PhyloNet with the best likelihood score has a single reticulation event in the ancestor of the red, yellow and green lineages (Figure 3.4), and the HyDe results for the same data recover a concentration of inferred hybrid taxa in the red and yellow clades (Figure 3.3). Interestingly, there are multiple possible parental pairs recovered for each of the hybrids inferred by HyDe, and the parental taxa are distributed across the phylogeny and are often from different geographic areas (Figure 3.3). Subsequent investigations of pairwise genetic distance revealed that taxa from the same clade are more similar to each other, and taxa from different clades exhibit greater genetic distance. If gene flow was ongoing among populations, we would expect hybrids to be more similar to their parental taxa. The observed pattern is inconsistent with this expectation.

Furthermore, we were able to detect the signature of hybridization in cultivated sweetpotato using PhyloNet analysis of the BLAST filtered data and in the HyDe analysis of both datasets. Austin (1988a) posited that cultivated sweetpotato was the result of hybridization between *I. trifida* and *I. triloba*. Interestingly, both the PhyloNet and HyDe results suggest that *I. ramosissima* may have provided a greater contribution to the *I. batatas* genome than any sampled accession of *I. trifida*. However, it is possible that, with deeper population level

sampling, an *I. trifida* lineage would have been identified as a parent of cultivated sweetpotato. Furthermore, the HyDe results suggest that *I. triloba* or *I. cordatotriloba* could have served as a potential parent of hexaploid, cultivated sweetpotato. Both filtered datasets analyzed in HyDe also identify both *I. ramosissima* and *I. triloba* as parents of sweetpotato. No prior hypotheses of the tetraploid origin of *I. tabascana* have been provided in the literature. Previous phylogenetic analyses suggest that *I. tabascana* is closely related to *I. trifida* and *I. batatas* (Jarret et al. 1992; Rajapakse et al. 2004). However, *I. tabascana* was recovered in the clade containing *I. triloba* in this analysis. HyDe analyses using the BLAST filtered data identified *I. lacunosa* and *I. splendor-sylvae* as parents of *I. tabascana*; however, the dataset filtered by read mapping identified *I. triloba* and *I. ramosissima* as the parental taxa. Furthermore, none of the PhyloNet runs identified hybridization in the history of *I. tabascana*. Therefore, these results are highly inconclusive and increased sampling will be necessary to fully resolve the full history of reticulation within the Batatas complex.

### *Conclusion*

Taken together, results presented here suggest that ancient hybridization and independent origins of polyploidy have played an important role in the evolutionary history of the Batatas complex. Furthermore, we provide evidence that both cultivated sweetpotato and tetraploid *I. tabascana* arose via hybridization. Ancient hybridization has played a major role in the evolution and diversification of many lineages, including humans (Durand et al. 2011; Sankaraman et al. 2015), fish (Schumer et al. 2016; Meier et al. 2017), and fungi (Marcet-Houben and Gabaldón 2015). Ancient hybridization has surely shaped the evolutionary history of the Batatas complex.

### **References**

Abel, W. E., and D. F. Austin. 1981. Introgressive hybridization between *Ipomoea trichocarpa* and *Ipomoea lacunosa* (Convolvulaceae). Bull. Torrey Bot. Club 108:231–239.

- Amborella Genome Project. 2013. The *Amborella* Genome and the Evolution of Flowering Plants. *Science*. 342:1241089.
- Arumuganathan, K., and E. D. Earle. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Report*. 9:208–218.
- Austin, D. F. 1988a. Nomenclatural Changes in the *Ipomoea batatas* Complex (Convolvulaceae). *Taxon* 37:184–185.
- Austin, D. F. 1978. The *Ipomoea batatas* complex. I. Taxonomy. *Bull. Torrey Bot. Club* 105:114–129.
- Austin, D. F. 1988b. The taxonomy evolution and genetic diversity of sweetpotatoes and related wild species. Pp. 27–59 in P. Gregory, ed. *Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources: Report of the 1st sweet potato planning conference 1987*. CIP, Lima, Peru.
- Austin, D. F., and Z. Huáman. 1996. A Synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon* 45:3–38.
- Bennetzen, J. L. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115:29–36.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537.
- Cao, Q., A. Zhang, D. Ma, H. Li, Q. Li, and P. Li. 2009. Novel interspecific hybridization between sweetpotato (*Ipomoea batatas* (L.) Lam.) and its two diploid wild relatives. *Euphytica* 169:345–352.
- Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17:540–552.
- Cervantes-Flores, J. C., G. C. Yenchó, A. Kriegner, K. V. Pecota, M. A. Faulk, R. O. M. Mwanga, and B. R. Sosinski. 2007. Development of a genetic linkage map and identification of homologous linkage groups in sweetpotato using multiple-dose AFLP markers. *Mol. Breed.* 21:511–532.
- Chifman, J., and L. Kubatko. 2014. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics* 30:3317–3324.
- Comer, J. R., W. B. Zomlefer, C. F. Barrett, D. W. Stevenson, K. Heyduk, and J. H. Leebens-Mack. 2016. Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). *Mol. Phylogenet. Evol.* 97:32–42. Elsevier Inc.
- Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24:332–40.

- Diaz, J., P. Schmiediche, and D. F. Austin. 1996. Polygon of crossability between eleven species of *Ipomoea* section *Batatas* (Convolvulaceae). *Euphytica* 88:189–200.
- Doyle, J. J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–15.
- Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, J. C. Pires, J. H. Leebens-Mack, and C. W. DePamphilis. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* 10:61.
- Duncan, T. M., and M. D. Rausher. 2013. Morphological and genetic differentiation and reproductive isolation among closely related taxa in the *Ipomoea* series *Batatas*. *Am. J. Bot.* 100:1–11.
- Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.
- Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution.* 63:1–19.
- Eserman, L. A., G. P. Tiley, R. L. Jarret, J. H. Leebens-Mack, and R. E. Miller. 2014. Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* 101:92–103.
- FAO. 2016. Food and Agriculture Organization of the United Nations.
- Felsenstein, J., and J. A. Cavender. 1987. Invariants of phylogenies in a simple case with discrete states. *J. Classif.* 4:57–71.
- Fisher, S., A. Barry, J. Abreu, B. Minie, J. Nolan, T. M. Delorey, G. Young, T. J. Fennell, A. Allen, L. Ambrogio, A. M. Berlin, B. Blumenstiel, K. Cibulskis, D. Friedrich, R. Johnson, F. Juhn, B. Reilly, R. Shammass, J. Stalker, S. M. Sykes, J. Thompson, J. Walsh, A. Zimmer, Z. Zwirko, S. Gabriel, R. Nicol, and C. Nusbaum. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* R1.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, D. Philip, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* 8:1–43.
- Heyduk, K., D. W. Trapnell, C. F. Barrett, and J. I. M. Leebens-Mack. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linn. Soc.* 117:106–120.

- Hirakawa, H., Y. Okada, H. Tabuchi, K. Shirasawa, A. Watanabe, H. Tsuruoka, C. Minami, S. Nakayama, S. Sasamoto, M. Kohara, Y. Kishida, T. Fujishiro, M. Kato, K. Nanri, A. Komaki, M. Yoshinaga, Y. Takahata, M. Tanaka, S. Tabata, and S. N. Isobe. 2015. Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. *DNA Res.* 1–9.
- Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254–67.
- Jarret, R. L., N. Gawel, and A. T. Whittemore. 1992. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *J. Am. Soc. Hortic. Sci.* 117:633–637.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Khoury, C. K., B. Heider, N. P. Castaneda-Alvarez, H. A. Achicanoy, C. C. Sosa, R. E. Miller, R. W. Scotland, J. R. I. Wood, G. Rossel, L. A. Eserman, R. L. Jarret, G. C. Yencho, V. Bernau, H. Juarez, S. Sotelo, S. de Haan, and P. C. Struik. 2015. Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series Batatas]. *Front. Plant Sci.*
- Kriegner, A., J. C. Cervantes, K. Burg, R. O. Mwanga, and D. P. Zhang. 2003. A Genetic Linkage Map of Sweetpotato (*Ipomoea batatas* (L.) Lam.) Based on AFLP Markers. *Mol. Breed.* 11:169–185.
- Kubatko, L., and J. Chifman. 2015. An Invariants-based Method for Efficient Identification of Hybrid Species From Large-scale Genomic Data. *bioRxiv* 34348.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–9.
- Lebot, V. 2009. Tropical root and tuber crops : cassava, sweet potato, yams and aroids. CABI, Wallingford, UK.
- Liu, L., L. Yu, and S. V. Edwards. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* 10:302.
- Loebenstein, G. 2009. Origin, Distribution and Economic Importance. Pp. 9–12 in G. Loebenstein and G. Thottappilly, eds. *The Sweetpotato*. Springer Netherlands.
- Loytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci.* 102:10557–10562.
- Löytynoja, A., and N. Goldman. 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science.* 320:1632–1635.

- Maddison, W. P. 1997. Gene Trees in Species Trees. *Syst. Biol.* 46:523–536.
- Marcet-Houben, M., and T. Gabaldón. 2015. Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13:1–26.
- McDonald, J. A., and D. F. Austin. 1990. Changes and Additions in *Ipomoea* section *Batatas* (Convolvulaceae). *Brittonia* 42:116–120.
- Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363.
- Mirarab, S., and T. Warnow. 2015. ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44–i52.
- Oracion, M. Z., K. Niwa, and I. Shiotani. 1990. Cytological analysis of tetraploid hybrids between sweet potato and diploid *Ipomoea trifida* (H. B. K.) Don. *Theor. Appl. Genet.* 80:617–624.
- Ozias-Akins, P., and R. L. Jarret. 1994. Nuclear DNA Content and Ploidy Levels in the Genus *Ipomoea*. *J. Soc. Hortic. Sci.* 119:110–115.
- Pamilo, P., and M. Nei. 1988. Relationships between Gene Trees and Species Trees. *Mol. Biol. Evol.* 5:568–583.
- Rajakapase, S., S. D. Nilmalgoda, M. Molnar, R. E. Ballard, D. F. Austin, and J. R. Bohac. 2004. Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear beta-amylase gene sequences. *Mol. Phylogenet. Evol.* 30:623–32.
- Refulio-Rodriguez, N. F., and R. G. Olmstead. 2014. Phylogeny of Lamiidae. *Am. J. Bot.* 101:287–299.
- Sankaraman, S., S. Mallick, M. Dannemann, K. Prufer, J. Kelso, S. Paabo, N. Patterson, and D. Reich. 2015. The landscape of Neandertal ancestry in present-day humans. *Nature* 507:354–357.
- Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* msw079.
- Schumer, M., R. Cui, D. L. Powell, G. G. Rosenthal, and P. Andolfatto. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25:2661–2679.
- Sharma, A. K., and P. C. Datta. 1958. Cytological investigations on the genus *Ipomoea* and its importance in the study of phylogeny. *Nucl.* 1:89–122.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

- Stephens, J. D., W. L. Rogers, K. Heyduk, J. M. Cruse-Sanders, R. O. Determann, T. C. Glenn, and R. L. Malmberg. 2015a. Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Mol. Phylogenet. Evol.* 85:76–87.
- Stephens, J. D., W. L. Rogers, C. M. Mason, L. A. Donovan, and R. L. Malmberg. 2015b. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. *Am. J. Bot.* 102:910–920.
- Storchova, H., R. Hrdlickova, J. Chrtek, M. Tetera, D. Fitze, and J. Fehrer. 2000. An Improved Method of DNA Isolation from Plants Collected in the Field and Conserved in Saturated NaCl/CTAB Solution. *Taxon* 49:79–84.
- Talavera, G., and J. Castresana. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* 56:564–577.
- Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.
- Tiley, G. P., L. A. Eserman, and R. E. Miller. *In prep.* Systematics, phylogenetics, and population genetic structure of sweet potato and its wild relatives: members of *Ipomoea* series *Batatas*.
- Ugent, D., and L. W. Peterson. 1988. Archaeological Remains of Potato and Sweet Potato in Peru. *Circ. Int. Potato Cent.* 16:1–10.
- Wen, D., Y. Yu, M. W. Hahn, and L. Nakhleh. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol. Ecol.* 2361–2372.
- Winslow, B. K. 2012. Interspecific Hybridization and Characterization of Variegation in Ornamental Sweetpotato (*Ipomoea batatas* (L.) Lam.). North Carolina State University.
- Wolcott, G. B. 1937. Chromosome numbers in the Convolvulaceae. *Am. Nat.* 71:190–192.
- Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proc. Natl. Acad. Sci.* 111.
- Yu, Y., and L. Nakhleh. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* 16:S10.

Table 3.1 – Accession information for individuals included in analyses in this study.

<b>Species</b>	<b>Accession</b>	<b>Locality</b>	<b>Accession source</b>
<i>I. batatas</i>	PI 561558	Tamaulipas, MX	USDA GRIN
<i>I. batatas</i>	PI 153655	cultivar Tinian	USDA GRIN
<i>I. cordatotriloba</i>	REM345	Florida, USA	J. S. Miller
<i>I. lacunosa</i>	REM844	Louisiana, USA	Frank Chalona
<i>I. ramosissima</i>	PI 552786	Bolivia	USDA GRIN
<i>I. setosa</i>	LAE74	cultivated	B&T World Seeds
<i>I. splendor- sylvae</i>	REM763	Costa Rica	R. Miller (CR08-17)
<i>I. tabascanana</i>	PI 518479	Tabasco, MX	USDA GRIN
<i>I. tenuissima</i>	PI 553012	Florida, USA	USDA GRIN
<i>I. tiliacea</i>	PI 165089	Puerto Rico	USDA GRIN
<i>I. trifida</i>	REM774	Costa Rica	R. E. Miller (CR08-43)
<i>I. trifida</i>	PI 618966	Michoacán, MX	USDA GRIN
<i>I. trifida</i>	ITR		Hirakawa et al. 2015
<i>I. trifida</i>	ITRk		Hirakawa et al. 2015
<i>I. trifida</i>	NSP306		<a href="http://sweetpotato.plantbiology.msu.edu/">http://sweetpotato.plantbiology.msu.edu/</a>
<i>I. triloba</i>	PI 536038	Veracruz, MX	USDA GRIN
<i>I. triloba</i>	PI 634796	Puerto Rico	USDA GRIN
<i>I. triloba</i>	NSP323		<a href="http://sweetpotato.plantbiology.msu.edu/">http://sweetpotato.plantbiology.msu.edu/</a>

Table 3.2 – Summary of genome size measurements, standard deviation from three flow cytometric measurements for each sample, and inferred ploidy for the accessions measured in this study.

Species	Accession	Genome size (pg/2C)	Standard Deviation	Inferred Ploidy
<i>I. batatas</i>	PI 561558	0.91	0.040	2x
<i>I. batatas</i>	PI 153655	2.69	0.044	6x
<i>I. cordatotriloba</i>	REM345	0.92	0.014	2x
<i>I. lacunosa</i>	REM844	0.90	0.011	2x
<i>I. ramosissima</i>	PI 552786	1.05	0.009	2x
<i>I. setosa</i>	LAE74	1.52	0.041	2x*
<i>I. splendor-sylvae</i>	REM763	1.06	0.003	2x
<i>I. tabascana</i>	PI 518479	1.96	0.040	4x
<i>I. tenuissima</i>	PI 553012			2x†
<i>I. tiliacea‡</i>	PI 165089			
<i>I. trifida</i>	REM774	0.94	0.060	2x
<i>I. trifida</i>	PI 618966	0.93	0.014	2x
<i>I. triloba</i>	PI 536038	0.87	0.030	2x
<i>I. triloba</i>	PI 634796	0.97	0.028	2x

\* Species determined to be diploid using chromosome counts by two previous studies (Wolcott 1937; Sharma and Datta 1958).

† Same accession determined to be diploid in Ozias-Akins and Jarret (1994).

‡ Plant died before leaf could be sampled for flow cytometry and was excluded from further analysis.

Table 3.3 – Summary of sequencing information for the accessions sequenced in this study.

Species	Accession	Raw Reads	Filtered reads	No. genes	Exon length	Intron length	Exon Coverage	Intron Coverage	Mean Gene Length
<i>I. batatas</i>	PI 561558	1789556	1789231	382	729.14	719.81	147.86	35.80	2455.78
<i>I. batatas</i>	PI 153655	371044	370959	256	534.84	590.30	19.05	7.58	1704.30
<i>I. cordatotriloba</i>	REM345	1033236	1033029	334	568.81	372.07	107.26	25.96	1648.40
<i>I. lacunosa</i>	REM844	1043820	1043623	231	298.61	260.45	57.14	16.68	951.78
<i>I. ramosissima</i>	PI 552786	3227800	3227181	180	415.49	511.13	184.37	53.60	1321.71
<i>I. setosa</i>	LAE74	464938	464863	367	793.19	1034.82	22.06	9.56	2795.48
<i>I. splendor-sylvae</i>	REM763	275692	275633	336	745.27	830.86	13.78	6.72	2380.65
<i>I. tabascana</i>	PI 518479	2305346	2304933	341	623.97	615.41	165.54	36.16	2103.39
<i>I. tenuissima</i>	PI 553012	1914722	1914385	376	745.09	745.52	172.93	40.30	2482.92
<i>I. tiliacea</i>	PI 165089	2387178	2386720	348	608.74	660.68	181.50	42.96	2140.98
<i>I. trifida</i>	REM774	2415222	2414773	238	489.73	506.07	175.72	44.21	1495.92
<i>I. trifida</i>	PI 618966	1373040	1372810	375	728.25	677.61	104.29	25.61	2353.64
<i>I. triloba</i>	PI 536038	441156	441076	352	828.32	1113.34	27.51	11.50	2926.76
<i>I. triloba</i>	PI 634796	1414288	1413976	326	516.85	445.05	132.12	30.60	1791.44

Table 3.4 – Alignment statistics for the five datasets used in analyses in this study.

<b>Taxa in dataset</b>	<b>Gene filtering</b>	<b>No. Individuals</b>	<b>No. Genes</b>	<b>Mean gene length</b>	<b>Total aligned length</b>	<b>PU sites<sup>†</sup></b>	<b>PI Sites<sup>‡</sup></b>	<b>Mean missing data</b>
<b>Diploids</b>	BLAST	15	366	1514.28	722,619	38,668	16,400	35%
<b>Dip + <i>I. batatas</i></b>	BLAST	16	366	1479.74	710,446	39,050	17,111	31%
<b>Dip + <i>I. batatas</i></b>	BLAST + read mapping	16	261	1344.29	456,369	25,621	11,150	34%
<b>Dip + <i>I. tabascana</i></b>	BLAST	16	366	1549.10	738,135	38,904	17,437	30%
<b>Dip + <i>I. tabascana</i></b>	BLAST + read mapping	16	351	1559.71	711,058	37,094	16,599	30%

<sup>†</sup> Parsimony uninformative sites

<sup>‡</sup> Parsimony informative sites

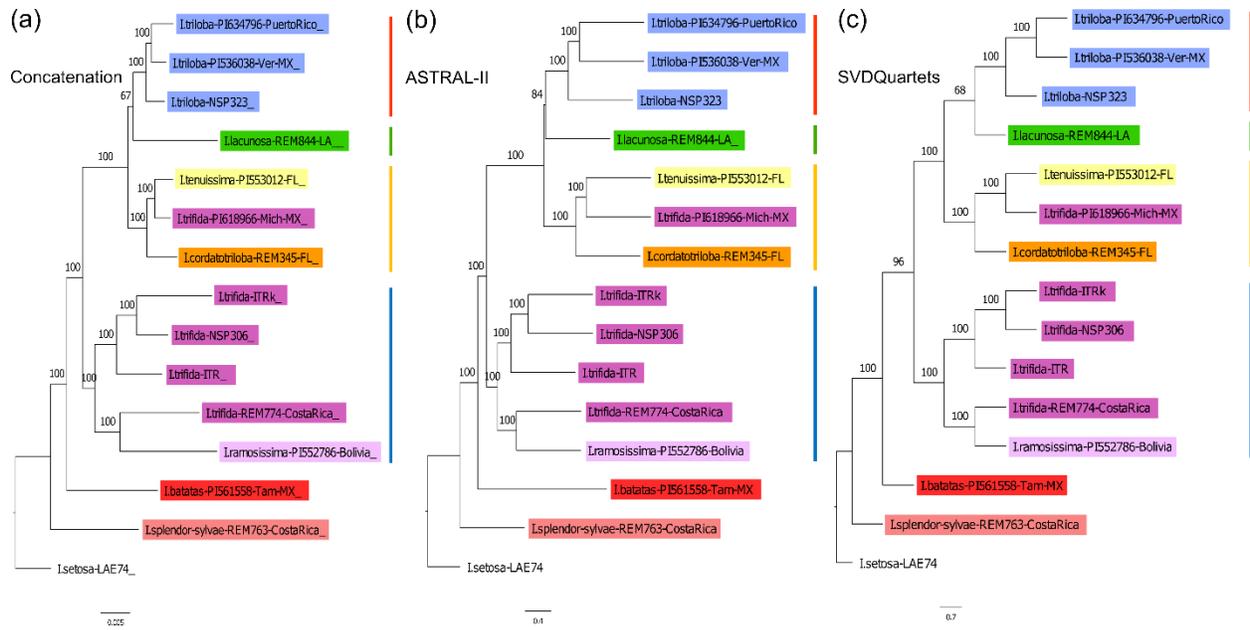


Figure 3.1 – Phylogenetic relationships among the diploid wild relatives of sweetpotato. Shown are the concatenated dataset analyzed with RAxML (a), the ASTRAL-II results (b), and the SVDQuartets tree (c). Each species is uniquely colored. To the right of each tree are bars denoting the three inferred major clades.

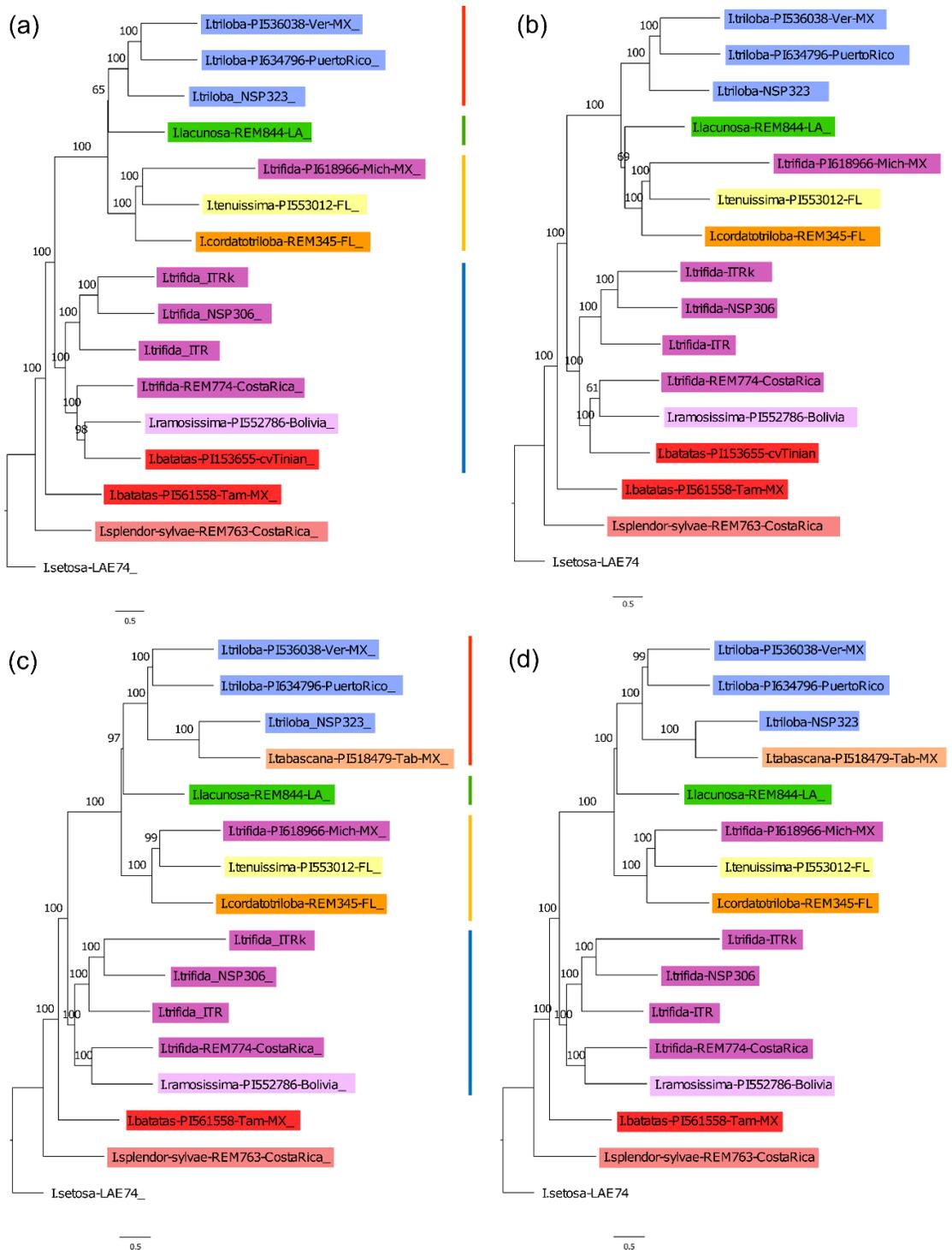


Figure 3.2 – Phylogenetic placement of hexaploid cultivated sweetpotato, cv. Tinian (a,b) and tetraploid *Ipomoea tabascana* (c,d). Panels a and c are trees estimated in ASTRAL-II from BLAST filtered datasets. Panels b and d are trees estimated in ASTRAL-II from datasets further filtered by examining read mapping. Species colors and clade bars correspond to coloring scheme in Figure 3.1.

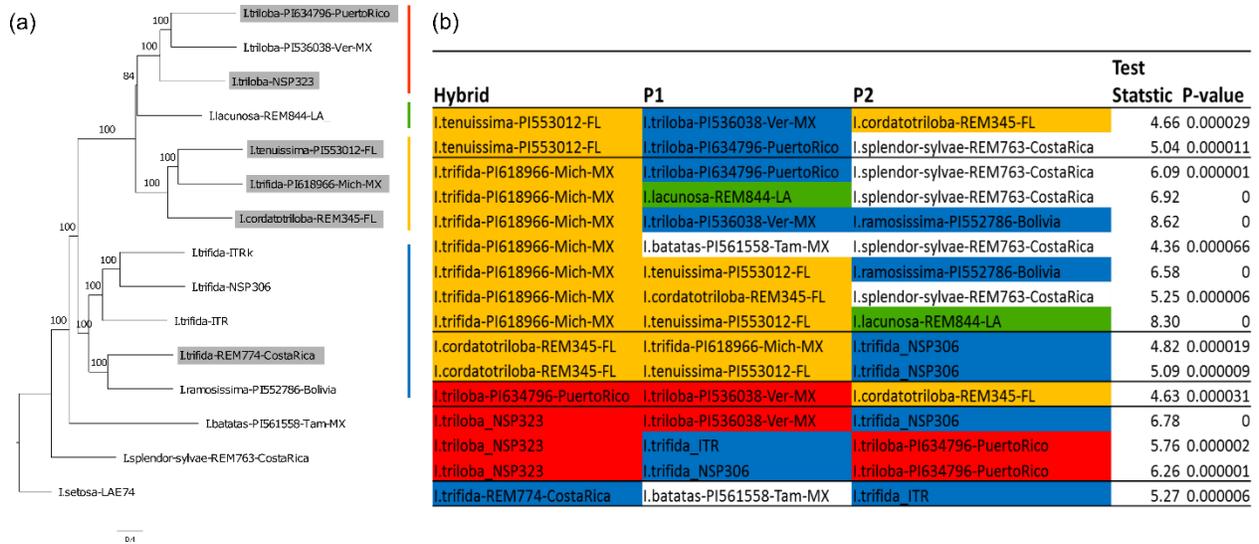
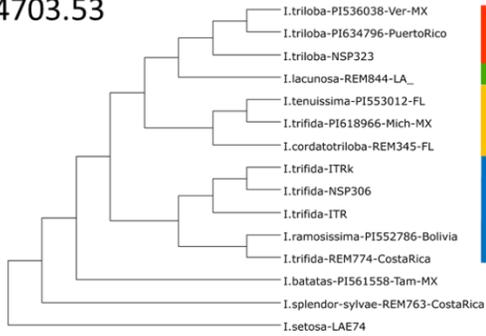
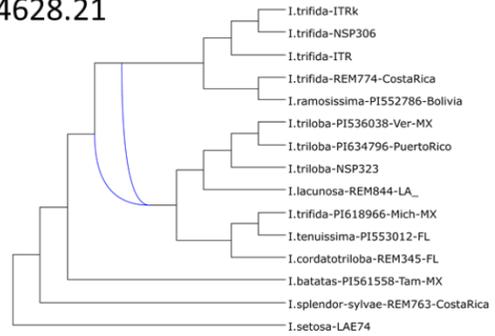


Figure 3.3 – Results of the HyDe analysis. (a) ASTRAL-II phylogeny showing species identified as hybrids in grey boxes. (b) Species identified as hybrids and inferred parental taxa. Also shown are the test statistic, p-value and bootstrap resampling support for the inferred hybridization scenarios from HyDe.

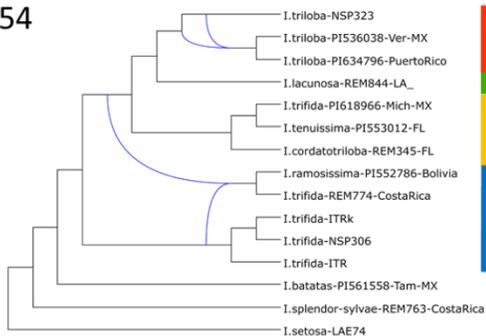
0 reticulations  
lnL = -4703.53



1 reticulation  
-4628.21



2 reticulations  
-4667.54



3 reticulations  
-4631.88

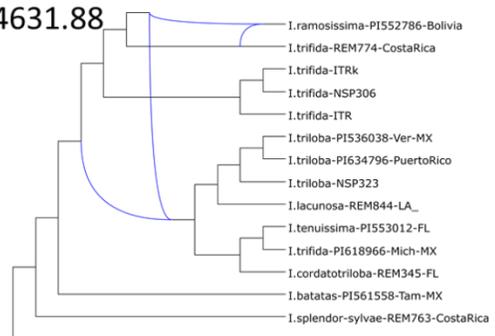


Figure 3.4 – Results from the PhyloNet analysis. Shown are the maximum likelihood estimates of reticulation under zero, one, two and three reticulation scenarios. Curved blue lines indicate inferred reticulation events from PhyloNet. Colored bars to the right of each network correspond to clade designations from Figure 3.1.

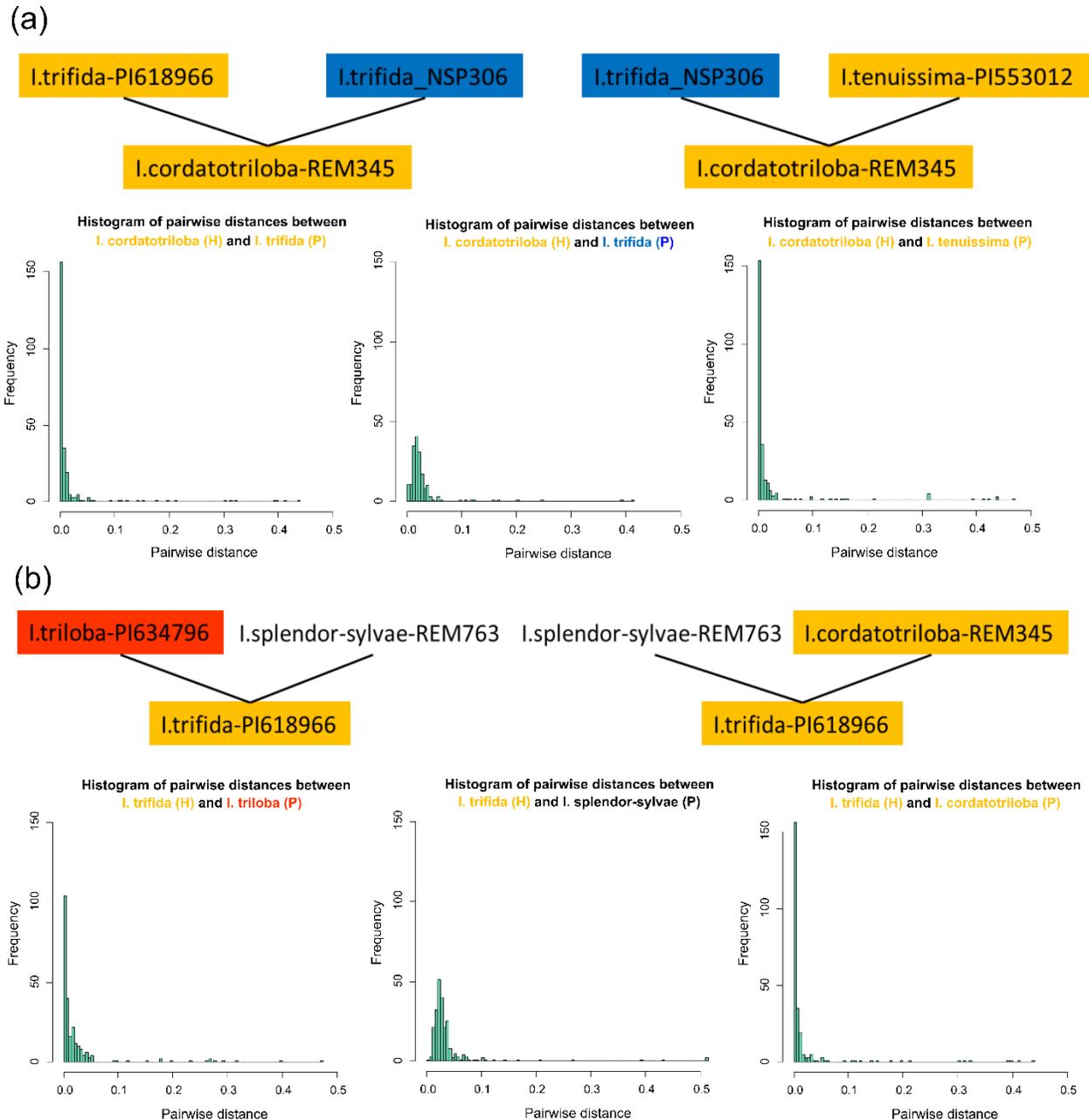


Figure 3.5 – Histograms depicting pairwise genetic distance (uncorrected p-distance) between taxa designated as hybrids and the inferred parental taxa from the HyDe analysis. Taxa are colored based on clade membership from Figure 3.1. (a) Genetic distance between *Ipomoea cordatotriloa* and the inferred parental taxa, *I. trifida* (PI618966), *I. trifida* (NSP306), and *I. tenuissima* (PI553012). (b) Genetic distances between *I. trifida* (PI618966) and inferred parental taxa *I. triloba* (PI634796), *I. splendor-sylvae* (REM763), and *I. cordatotriloa* (REM345).

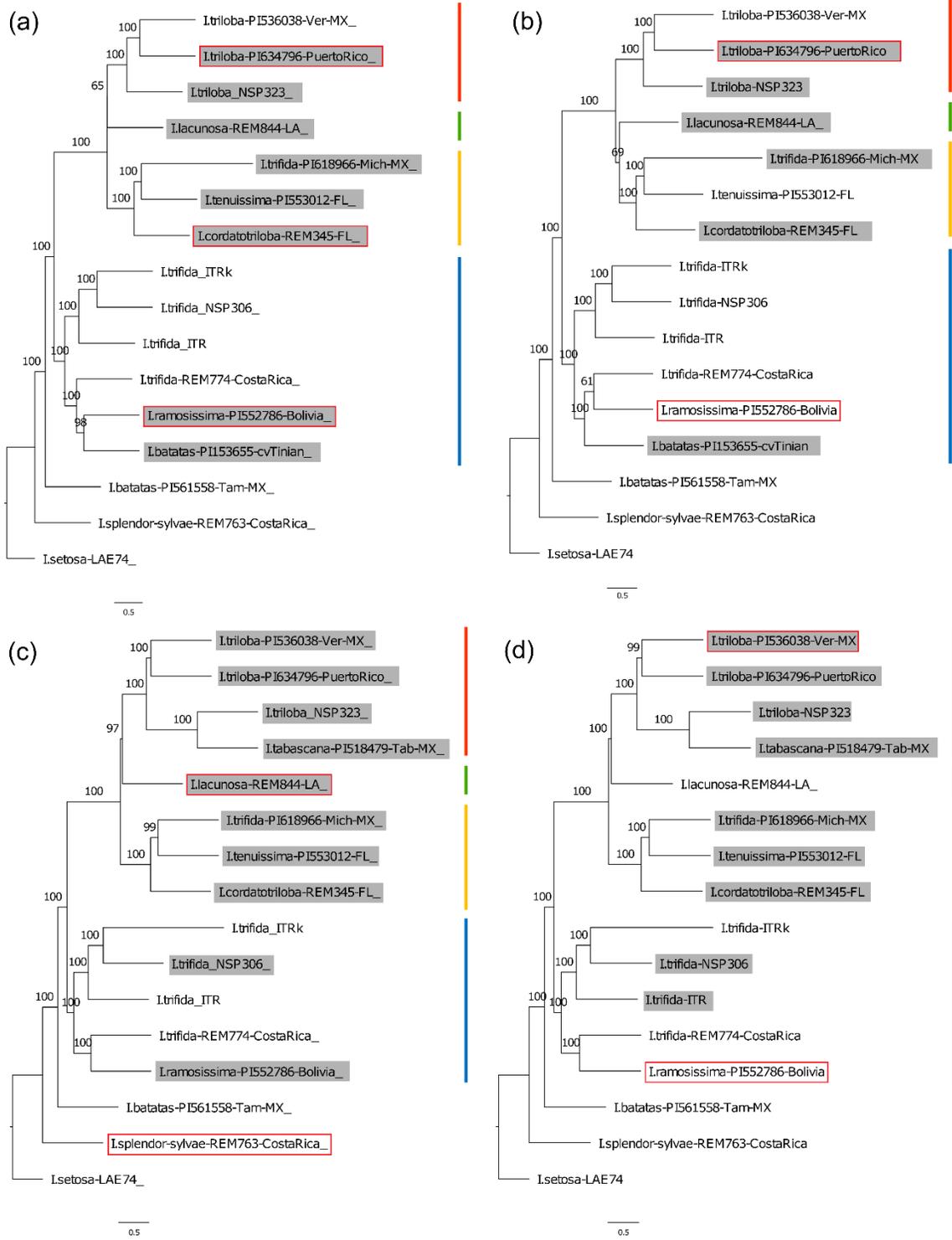


Figure 3.6 – Results of the HyDe analysis which included hexaploid *Ipomoea batatas* (a,b) and *I. tabascana* (c,d). Results from datasets filtered by BLAST only are shown in (a) and (c), and datasets filtered by read mapping are displayed in (b) and (d). Shown are ASTRAL-II trees with species identified as hybrids in grey boxes. Inferred parents of the hybrid polyploid taxa are bound by red boxes.

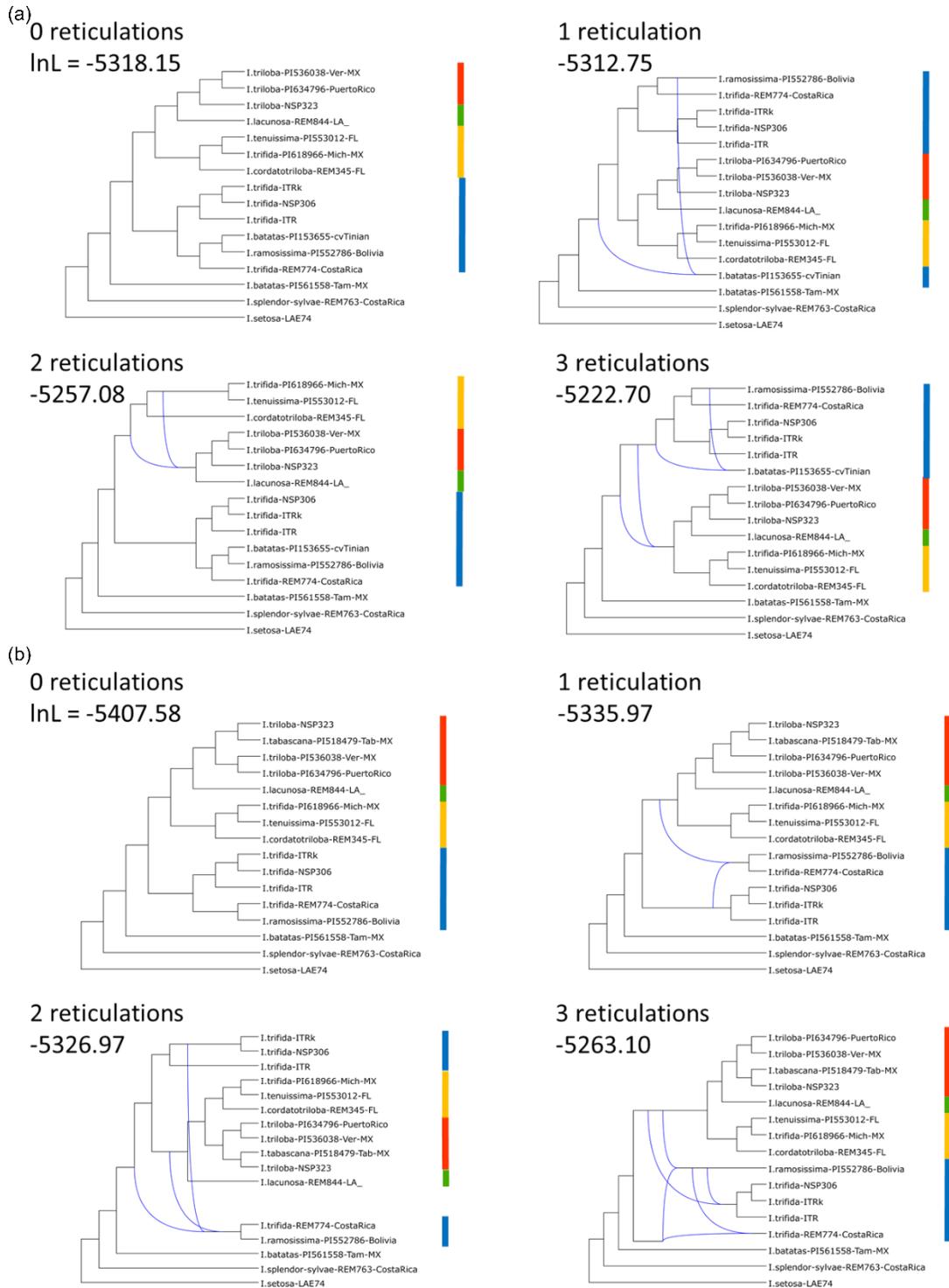


Figure 3.7 – Results from the PhyloNet analyses which included *Ipomoea batatas* (a) and *I. tabascana* (b). Shown are the maximum likelihood estimates of reticulation under zero, one, two and three reticulation scenarios. Curved blue lines indicate inferred reticulation events from PhyloNet. Colored bars to the right of each network correspond to clade designations from Figure 3.1.

CHAPTER IV:

EVOLUTION OF POLYPLOIDY AND STORAGE ROOTS IN SWEETPOTATO AND ITS WILD RELATIVES

(*IPOMOEA* SERIES *BATATAS* (CHOISY) D. F. AUSTIN)<sup>3</sup>

<sup>3</sup>Eserman, Lauren A., Jim H. Leebens-Mack. To be submitted to *Systematic Biology*.

## **Abstract**

Several previous studies have noted a correlation between various plant traits, such as growth rate, cell size, and geophytic habit, with genome size, suggesting more slowly growing plants with larger cells and underground storage organs tend to have larger genomes. Previous studies have noted a potential association between genome size and storage root formation within the Batatas complex of the morning glory family (Convolvulaceae), which includes sweetpotato and its wild relatives. Using data from target sequence capture, we identified four major clades, corresponding to the same four identified in Chapter 3. Further, we find that most individuals sampled were diploid, and there were at least two independent origins of polyploidy, with at least two rounds of polyploidy in the clade containing *Ipomoea batatas*. Cultivated sweetpotato as well as tetraploid wild *I. batatas* individuals were closely related to *I. trifida* in phylogenetic analyses. We further tested for hybridization with HyDe and PhyloNet and found evidence for a single ancient hybridization event in the Batatas complex. Though we found a significant correlation between genome size and root traits, the correlation was weak suggesting there are other factors contributing to the observed variation in root traits. Ancient hybridization may have played a role in introgression of loci controlling root traits or key taxonomic characters. Future examinations of hybridization in the Batatas complex should consider the impact on the evolution of phenotypic traits.

## **Introduction**

Understanding the mechanisms driving phenotypic diversification is a fundamental goal of evolutionary biology. Numerous studies have documented a positive association between plant genome size and ploidy level with quantitative phenotypic traits. This correlation has been found at multiple levels of cellular organization. Specifically, genome size is positively

correlated with cell size (Edwards and Endrizzi 1975), seed size (Knight and Ackerly 2002), stomatal guard cell density (Knight et al. 2005; Knight and Beaulieu 2008) but negatively correlated with photosynthetic rate and growth rate (Knight et al. 2005). In addition, studies have also noted that geophytes – species with underground storage organs – tend to have larger genomes than non-geophytes (Grime and Mowforth 1985; Veselý et al. 2012). Furthermore, these correlations hold for increases in genome size due to polyploidy and other factors such as retrotransposon proliferation (Knight and Beaulieu 2008). The number of documented correlations are many; however, few studies have examined associations between ploidy level and specific root traits.

Plant roots display a variety of complex phenotypes and can be modified in a myriad of ways. A typical angiosperm root consists of a primary taproot with laterals extending outward from the taproot. Occasionally, the taproot and/or one or more of the lateral roots will enlarge to form a starch storage organ. Starch-rich storage roots exhibit an accumulation of starch-filled cells comprising the anomalous cambium, a tissue type associated specifically with starch storage (Artschwager 1924; Wilson and Lowe 1973; Lowe and Wilson 1974a, 1974b). Anomalous cambium proliferates as a storage root swells and accumulates starch (Noh et al. 2013). Given the documented associations between genome, cell and organ sizes, we aim to test whether species within the Batatas complex that form storage roots have larger genomes than closely related species lacking storage roots.

The morning glory clade containing sweetpotato [*Ipomoea batatas* (L.) Lam.] and its wild relatives (hereafter, the Batatas complex) represents an ideal system for investigating the relationships between genome size, root cell size, root development and storage root formation. The Batatas complex consists of 14 named species with ploidy levels ranging from 2x to 6x

(Ozias-Akins and Jarret 1994; Diaz et al. 1996; Roullier et al. 2013b). It has been hypothesized that polyploidy may have played a role in the development of large storage roots in sweetpotato (Reddy et al. 2007). This hypothesis is especially compelling, as cultivated sweetpotato is hexaploid and makes large storage roots, while most diploid wild relatives are not known to produce storage roots (McDonald 1994; Ozias-Akins and Jarret 1994).

The evolutionary history of the Batatas complex has been poorly characterized in the past. Three studies have examined the phylogeny specifically of the Batatas complex (Austin 1988b; Jarret et al. 1992; Rajapakse et al. 2004). These studies generally examine a single individual per species and thus cannot test the monophyly of species within the Batatas complex. Further, while *Ipomoea batatas* has been implicated as allopolyploid, none of these previously published phylogenetic analyses of the Batatas complex have accounted for possible reticulations within the evolutionary history of the group. Previous work has also failed to account for gene tree – species tree discordance due to retention of ancestral allelic variation between speciation events, or incomplete lineage sorting (ILS). Understanding the role of reticulate evolution and ILS has important implications for inferences about the genetic basis of phenotypic evolution (Hahn and Nakhleh 2015).

In this study, we seek to understand whether increases in genome size due to polyploidy have played a role in the evolution of storage roots. Critical to this examination is a robust species phylogeny for the Batatas complex as well as an understanding of the impacts that ILS, polyploidy and reticulations have on genome evolution in this group. Therefore, the primary goals are to (1) reconstruct a phylogeny of the Batatas complex sampling multiple individuals per species when possible, (2) characterize genome size and root traits, (3) test for a correlation between genome size and root traits while accounting for phylogeny, and (4) understand how

hybridization has shaped the evolution of species and phenotypic traits in the Batatas complex. Contrary to our expectations, we find that most individuals sampled from the Batatas complex are diploid, and storage root formation has evolved independently of any increases in genome size. Furthermore, ancient hybridization has certainly shaped the evolutionary history of the Batatas complex.

## **Methods**

### *Taxon sampling*

Individuals were selected to represent the breadth of morphological and geographic diversity in the Batatas complex (Table 4.1). Multiple individuals were sampled for 8 species. One accession was used for *I. cynanchifolia*, *I. tabascanana*, and *I. tenuissima*. Two accessions of *I. splendor-sylvae* were sampled, but one accession failed to germinate. Seeds or cuttings were obtained from a variety of sources, primarily from USDA-GRIN and the research collections of Drs. Rick Miller and Michael T. Clegg. Representative individuals were sampled from all but two named species in the Batatas complex, *I. littoralis* and *I. tiliacea*. In total, sixty-four accessions were sequenced.

### *Plant material*

Four seeds or cuttings of each accession were grown under controlled conditions in the UGA greenhouses. One seed or cutting, hereafter referred to as the reference plant, was planted and allowed to grow to flowering. Leaves from this accession were used for genome size measurements and DNA isolation. The remaining three seeds or cuttings, hereafter referred to as experimental plants, were grown for measurement of root traits. Experimental plants were grown for four weeks after germination or rooting to measure variation in the early stages of root development.

### *Genome size measurements*

Genome size measurements and ploidy inference were carried out as described in Chapter 3 on 59 of the 64 accessions sequenced in this study (Table 4.2). Four accessions measured in the previous study (Ozias-Akins and Jarret 1994) were repeated to determine accurate ploidy estimation across experiments (Table 4.3). Genome sizes were measured across two years, and four accessions were repeated to confirm measurement accuracy across years (Table 4.4). Nuclear DNA content and ploidy are highly positively correlated in these species (Ozias-Akins and Jarret 1994); therefore, genome size served as a proxy for ploidy level.

### *Root tissue sampling*

Experimental plants were grown for four weeks after seed germination or slip planting. Seeds or slips were randomized and planted 50 at a time over the course of four weeks for a total of seven batches of plants; staggered planting allowed for a reasonable number of plants for washing and root sampling on a given day. Four-week-old plants were removed from pots, most dirt was gently removed from roots by hand, and roots were then carefully washed with tap water to remove remaining soil particles.

### *Measurement of root traits*

Images of roots were then taken on a 2' x 2' sanded plywood board painted with black chalkboard paint (Bucksch et al. 2014). Images were taken using a 12.2 megapixel Samsung camera mounted horizontally on a camera stand, following Bucksch et al. (2014). Measurements of root width were made using the morphometric software package tpsDig2 (Rohlf 2010). Images were calibrated with a ruler which was placed next to each root during imaging. Three measurements were taken of the taproot: 1) the crest of the taproot, 2) where the 4th lateral root emerges and 3) where the 10th lateral root emerges. These regions were previously identified to

have a high degree of variation between individuals that do and do not form storage roots (Eserman et al. *in prep*). Width measurements of the smallest and largest lateral roots were also recorded.

After imaging, root tissue was dried at 50°C for 24-36 hours in a drying oven. Total starch content was measured separately on fine roots and taproots using the following starch measurement protocol, modified from (Hansen and Moller 1975; Oren et al. 1988; Zeeman et al. 1998); 0.20 grams of dried root was ground in a mortar and pestle and washed once with room temperature acetone to remove any tannins and chlorophyll. Anthocyanins were then extracted with 80% ethanol and discarded. The dried and anthocyanin-extracted tissue was then heated in a 1% hydrochloric acid solution to solubilize starch. Root tissue was pelleted by centrifugation. An 1:40 dilution of 5% Lugol's iodine (Fisher Scientific, Waltham, MA) was added to the crude starch extract, and optical density was measured at 580 nm using a Turner SP-830 spectrophotometer. Optical density was converted to starch concentration with a standard curve, made using potato starch dissolved in 1% hydrochloric acid solution.

#### *Library preparation and sequencing*

DNA isolation, library preparation, sequence capture and sequencing were carried out using methods described in Chapter 3.

#### *Sequence assembly*

Sequence assembly was carried out generally following the reads2trees pipeline in a similar manner to methods described in Chapter 3 (Heyduk et al. 2015). Reads were first sorted by barcode. Adapter and barcode sequences were trimmed from reads using Trimmomatic (Bolger et al. 2014). In addition, reads less than 40 bp in length after adapter and barcode trimming were removed using Trimmomatic. Trinity version 2.0.6 (Haas et al. 2013) was used to

*de novo* assemble reads into contigs. BLAST (Camacho et al. 2009) was then used to match assembled contigs to the reference exon set used to design baits. For each sample, any instance where more than one contig matched the reference exon set, the contigs were removed in order to avoid possible problems with misspecification of orthology. The BLAST filtered genes assembled in Chapter 3 were also added to these data for a total of 77 accessions. The samples were sorted into two datasets: one containing only diploid taxa and the second containing diploid and polyploid taxa. PRANK (Loytynoja and Goldman 2005; Löytynoja and Goldman 2008) was used to align genes, and Gblocks (Castresana 2000; Talavera and Castresana 2007) was applied to remove regions with poorly supported alignments.

#### *Phylogenetic methods*

Gene trees were estimated in RAxML, and branch support was assessed for each gene tree using the rapid bootstrapping algorithm (Stamatakis 2014). RAxML was also used to estimate a phylogeny given a concatenated alignment of all genes for both the diploid and polyploid + diploid datasets. To account for the possible influence of incomplete lineage sorting (ILS), trees were constructed using both ASTRAL-II (Mirarab and Warnow 2015) and SVDQuartets (Chifman and Kubatko 2014), which have both been shown to converge on the correct topology while accounting for ILS. In this analysis, maximum likelihood gene trees from RAxML were used as input in ASTRAL-II, and the multilocus concatenated alignment was used in SVDQuartets. In both analyses, branch support was evaluated using 100 bootstrap replicates. *Ipomoea setosa* was used as the outgroup for all phylogenetic analyses. In individual gene alignments where *I. setosa* was not present, *I. sepacuitensis*, a sister species to *I. setosa* (McDonald et al. 2011), was used as the outgroup.

### *Inferring hybridization*

HyDe was used to infer potential hybrid and parental taxa while also accounting for ILS (Kubatko and Chifman 2015). The concatenated multilocus gene alignment was used as the input for HyDe. This test assesses the frequency of site patterns in a set of three ingroup and one outgroup individuals. Asymmetries in the site patterns are indicative of hybridization, whereas equal frequencies of a particular site pattern are indicative of ILS (Kubatko and Chifman 2015). HyDe can only use a single outgroup taxon; therefore, *Ipomoea sepacuitensis* was removed from the alignment and *I. setosa* was used as the outgroup for HyDe analyses. The p-value for a significant inference of a particular hybridization scenario was adjusted for multiple comparisons using the Bonferroni correction suggested by Kubatko and Chifman (2015). A p-value of 3.8566E-7 was applied to the diploid dataset, and a p-value of 3.0428E-7 was used for the dataset that included polyploid taxa. At least 150 sites were required to share a site pattern for the test of hybridization to occur. To further examine the stability of the HyDe results, we re-ran the HyDe analysis requiring 50 and 10 sites to share a site pattern.

In the previous chapter, the maximum pseudolikelihood method applied in PhyloNet (Than et al. 2008; Yu and Nakhleh 2015) was used to test for hybridization. However, PhyloNet analyses are computationally intensive and are limited by the number of individuals included in the analysis (Yu et al. 2014). Therefore, we are not currently able to use PhyloNet to test for reticulations on this phylogeny. In order to accommodate this limitation while testing the robustness of the inference of ancient reticulation in the smaller analyses described in Chapter 3, we randomly generated five datasets with the same number of taxa that were included in the previous PhyloNet analyses. In all five datasets, we included *I. setosa*, *I. batatas*, and *I. splendora-sylvae* but randomly sampled individuals from the full dataset which were placed in the red,

yellow, green and blue clades described in Chapter 3. This allowed us to test whether, given the same phylogenetic structure but different sets of taxa, we inferred the same hybridization events identified in Chapter 3. For each of the five datasets, alignments were re-estimated in Prank (Loytynoja and Goldman 2005; Löytynoja and Goldman 2008) and cleaned using Gblocks (Castresana 2000; Talavera and Castresana 2007). Gene trees were estimated in RAxML (Stamatakis 2014), and the maximum likelihood gene trees were used as input for the PhyloNet analyses. In PhyloNet, we estimated a phylogenetic network given the set of gene trees under zero, one, two, and three reticulation scenarios using the maximum pseudolikelihood approach (Yu and Nakhleh 2015).

#### *Phylogenetic comparative methods*

We employed phylogenetic comparative methods treating traits as both discrete and continuous characters. We first examined the evolutionary history of ploidy on the Batatas complex phylogeny using the inferred ploidy data as a discrete trait using the topology recovered in the concatenation analysis. Although we recognize that in many cases the topology of a tree generated from analysis of a concatenated alignment likely does not reflect the true species tree, the concatenation tree was the most logical choice to use in this scenario for two reasons. First, SVDQuartets does not estimate any branch lengths, and ASTRAL-II does not estimate terminal branch lengths, and branch lengths are necessary for maximum likelihood analysis of character state evolution (Pagel 1994, 1999). Second, the topology of the concatenation tree did not differ greatly from the ASTRAL-II and SVDQuartets trees. We applied a maximum likelihood approach to reconstruct ancestral character states for ploidy. We applied the Mk1 model of discrete trait evolution, which imposes equal transition rates among all character states (Pagel

1994; Mooers and Schluter 1999), in the phytools R package using the “ace” function (Revell 2012) in R version 3.3.3.

Trait value distributions were tested for normality using the Shapiro-Wilk’s test for normality (Shapiro and Francia 1972) using the shapiro.test function in R version 3.3.3. In this test, the null hypothesis is that the data fit a normal distribution, and a significant p-value indicates that the data deviate significantly from a normal distribution. Genome size, width of the taproot at the fourth lateral root, and thickness of the thinnest lateral root, and starch concentration in both fine and thicker roots were all found to be significantly different from the null hypothesis. Therefore, these data were transformed to fit a normal distribution. Genome size exhibited an extreme right skew, so a log-transformation was applied. Taproot width at the fourth lateral root and thickness of the thinnest lateral root both showed a slight left skew, and values were square-transformed. Finally, starch concentration in both fine and thicker roots exhibited a slight skew to the right, and a square root transformation was applied.

We then calculated two measures of phylogenetic signal for root traits and genome size, Blomberg’s K (Blomberg et al. 2003) and Pagel’s lambda (Pagel 1999). These two metrics quantify phylogenetic signal in two different ways. Blomberg’s K divides the among species variance in a trait by the variance in contrasts for that trait (Blomberg et al. 2003); whereas, Pagel’s lambda compares trait correlations among species to correlations that would be expected under a Brownian motion model (Pagel 1999). For both measures of phylogenetic signal, a value close to zero implies that trait values are distributed randomly with respect to the phylogeny and a value close to one suggests that trait values are perfectly correlated with phylogeny as would be expected under a model of Brownian motion (Münkemüller et al. 2012). We used the “phylosig” function in phytools (Revell 2012) in R version 3.3.3 to test for phylogenetic signal. For root

traits, we included the mean of three biological replicates for each accession, and the mean of three technical replicates was incorporated for genome size. Fifteen accessions did not have the required three biological replicates for root traits due to low seed germination, so our comparative analyses did not include all of the accessions included in the phylogenetic analysis.

The ultimate goal was to test for correlated evolution of genome size and root traits. It would have been possible to test for correlations of genome size and root traits for each root measurement separately, but the root size trait measurements are likely highly correlated. Therefore, we employed a commonly used approach whereby variation in continuous traits is collapsed using a phylogenetically corrected principal components analysis (pPCA) (Revell 2009). The pPCA was carried out using the “`phyl.pca`” function in `phytools` in R version 3.3.3 (Revell 2012) using the following root traits: (1) width of the top of the primary root, (2) width of the primary root where the fourth lateral root emerged, (3) width of the primary root where the tenth lateral root emerged, (4) thickness of the thickest lateral root, (5) thickness of the thinnest lateral root, (6) primary root starch concentration, and (7) lateral root starch concentration.

We then calculated phylogenetic independent contrasts for the pPC axes and genome size using the “`pic`” function in `phytools` under R version 3.3.3. There was a large amount of variation in measurements of starch concentration, so the contrasts were calculated for pPCA axes with and without starch concentration. The first two pPCA axes with and without starch concentration were used in a general linear model with log-transformed genome size to test for a correlation between genome size and root traits.

## Results

### *Genome size measurements*

Genome size was measured on fifty-nine individuals in this experiment and ranged from  $0.81 \pm 0.025$  to  $2.63 \pm 0.021$  pg/2C nucleus (Table 4.2). The mean genome size across all individuals was 1.05 pg/2C nucleus. Cultivated sweetpotato (*I. batatas* cv. Beauregard) had the largest genome size with a value of 2.63 pg/2C nucleus. Four accessions measured in a previous study (Ozias-Akins and Jarret 1994) were measured here (Table 4.3). Overall, genome size measurements were much higher in the previous compared to the present study. In three of the four accessions, inferred ploidy was the same. However, *I. cordatotriloba* PI518495 was inferred to be tetraploid in the previous study, but it was inferred to be diploid in this experiment. Genome size measurements were all performed in the same lab but were done across two years. Therefore, four accessions were repeated across years to ensure measurement accuracy. In all four cases, measurements across years were within the range of standard error (Table 4.4).

### *Sequencing results*

In total, sixty-four libraries were sequenced for this project (Table 4.5). One sample, *Ipomoea batatas* REM356, had very low read numbers (3,528 reads), so this sample was removed from further analysis due to low sequencing depth. Three samples, *I. cordatotriloba* PI518494, *I. trifida* E/Pau27, and *I. trifida* PI543830, were removed from further analysis because the plant died before genome size could be measured. After removal of these samples, mean gene length ranged from 565.93 bp in *I. batatas* cv. Beauregard to 1723.35 bp in *I. triloba* PI634795 (Table 4.5). Exon length ranged from 196.03 bp in *I. batatas* cv. Beauregard to 708.02 bp in *I. grandifolia* PI561550. Again, *I. batatas* cv. Beauregard had the smallest intron length (366.70 bp), and the longest intron length was recovered in *I. triloba* PI634795 (1041.45 bp).

Overall, intron coverage was lower than exon coverage, and these measures varied widely among libraries. Exon coverage ranged from 94.45x in *I. batatas* cv. Jewel to 3264.88x in *I.* unknown LAE-FL19. Intron coverage was lowest in *I. leucantha* PI518481 (26.71x) and was highest in *I. triloba* PI540710 (1291.24x).

### *Molecular data matrices*

The data matrices presented here include the 60 individuals along with the 17 individuals used in phylogenetic analyses in Chapter 4. The concatenated gene alignment for only the diploid taxa contained 71 individuals and 244 genes (Table 4.6). The diploid species data matrix was 505,689 bp in length and contained 44,297 variable but parsimony uninformative sites, where the minor allele is found in only one sample, and 25,335 parsimony informative sites, where the minor allele is found in two or more samples. The dataset which included both diploid and polyploid individuals contained 77 taxa and 244 genes. The diploid + polyploid data matrix was 490,520 bp in length. The polyploid + diploid dataset contained 43,600 sites which were variable but parsimony uninformative and 25,849 parsimony informative sites (Table 4.6).

### *Phylogenetic results*

#### Diploid taxa

In trees reconstructed using only diploid individuals, the Batatas complex was recovered as monophyletic with 100% bootstrap support (Figure 4.1, S4.1, S4.2). *Ipomoea splendor-sylvae* was recovered as sister to the rest of the Batatas complex in all trees with high bootstrap support (>97%). In the trees including only diploid taxa, five separately evolving lineages were identified in the ASTRAL-II and concatenation trees, and four main lineages were recovered in the SVDQuartets tree (Figure 4.1, S4.1, S4.2). In all three trees, the red and green lineages were sister to one another with bootstrap support >81%. In all three trees, the lineage containing

members of the red, green, and yellow clades was sister to the blue clade. The red clade was found to contain primarily *I. triloba* and *I. leucantha* individuals (Figure 4.1, S4.1, S4.2). The green lineage contained a mixture of species, such as *I. austinii*, *I. cynanchifolia*, and *I. grandifolia* (Figure 4.1, S4.1, S4.2). The yellow clade was primarily made up of *I. lacunosa* and *I. cordatotriloba* individuals. The blue clade contained mostly *I. trifida* accessions. Multiple individuals were included for eight species. Seven of these eight species were paraphyletic, and *I. austinii* was the only species recovered as monophyletic. *Ipomoea trifida* was primarily concentrated in the blue clade but was distributed across the red and yellow clades as well.

#### Placement of polyploid taxa

When polyploid taxa were included in the phylogenetic analyses, the resulting topologies were similar to what was recovered when analyzing only the diploid taxa. The same five major clades were recovered in the ASTRAL-II and concatenation trees (Figure 4.2, S4.4). Three hexaploid and two tetraploid *Ipomoea batatas* individuals were included in this tree. These five polyploid *I. batatas* individuals were recovered in a subclade of the blue clade which also contained *I. ramosissima* PI552786 and *I. trifida* REM774. However, this subclade was recovered with low bootstrap support (ASTRAL-II = 20%, SVDQuartets = 67%, concatenation = 78%). Furthermore, the relationships among these individuals could not be recovered with certainty, as many branches in this subclade also had low bootstrap support (<70% in all trees). The diploid *I. batatas* diverged prior to the diversification of the red, yellow, blue and green clades. The polyploid *I. batatas* individuals diverged from a subset of *I. trifida* individuals. The tetraploid *I. tabascanana* was placed within the red clade, which is dominated by *I. triloba*. This relationship was supported with bootstrap support of 100% in all phylogenetic analyses including polyploids.

### *Genome size evolution*

We estimated the evolution of ploidy level as a discrete trait using the Mk1 model. In total, there were three tetraploid and three hexaploid individuals, and the rest of the samples were diploid. It appears that there were two independent origins of polyploidy in the Batatas complex. One in the red clade giving rise to the tetraploid *I. tabascana* lineage (Figure 4.3). The other is in the blue clade giving rise to a subclade including the polyploid *I. batatas* lineages (Figure 4.3). There was very little resolution among the individuals in this clade; therefore it is difficult to pinpoint the exact timing of the whole genome duplication events. However, there had to be at least two rounds of polyploidization, at least one giving rise to tetraploids and a second round giving rise to hexaploid *I. batatas* (Figure 4.3).

### *Inference of hybridization*

We applied the software HyDe (Kubatko and Chifman 2015) to infer hybrids and their potential parental taxa. The HyDe results for the dataset including only diploid taxa resulted in a total of forty-seven identified hybrid taxa, regardless of whether we required 150, 50 or 10 sites to share a site pattern. Similar to results found in Chapter 4, the inferred hybrids were concentrated in the lineage containing the red, yellow and, green clades (Figure 4.4). There was no obvious pattern with regard to the identified parental taxa. The putative parents of the hybrid taxa were distributed across the phylogeny and were often from very distant geographic regions. A similar pattern was recovered for the dataset including diploids and polyploids, where inferred hybrids were concentrated in the lineage containing the red, yellow, and green clades, and the same set of hybrids were identified in analyses requiring 150, 50, and 10 sites to share a site pattern (Figure 4.5). Again, parental taxa were sampled from across the entire phylogeny and were from different geographic areas. In contrast, a single individual from the blue clade, *I.*

*leucantha* PI518481, was identified as a hybrid in the analysis containing polyploid taxa; however, this individual is diploid (Figure 4.5). Similar to results from Chapter 4, *I. tabascana* was identified as having hybrid ancestry in this analysis. Surprisingly, however, none of the polyploid *I. batatas* accessions were identified as hybrid in these analyses, regardless of the number of sites used to test for hybrid ancestry. Two of the five polyploid *I. batatas* samples were very nearly significant; hexaploid *I. batatas* PI 153655 (cultivar Tinian) and tetraploid *I. batatas* PI518474 were identified as hybrid with a p-value  $< 9.0E-7$ , when the p-value for significance using the Bonferroni method was  $3.70E-7$ . In all of these nearly significant cases, either hexaploid or tetraploid *I. batatas* was identified as a hybrid with one parent coming from the blue clade and *I. cordatotriloba* (REM345) as the other parent. Other results with p-value  $< 9.0E-6$  identify hexaploid or tetraploid *I. batatas* as a hybrid with one parent coming from the blue clade and the other sampled from either the red or yellow clade.

The maximum pseudo-likelihood analysis in PhyloNet was also run on five randomized datasets including taxa from this larger experiment. Analysis of the first random sample (Figure 4.6a) recovered three reticulations, two of which were in the ancestor of the lineage containing the red, green, and yellow clades, as seen in the analyses described in Chapter 3. The third reticulation was in the ancestor of only the red clade. The second analysis (Figure 4.6b) also showed three reticulations. One was in the ancestry of *I. lacunosa* REM844, and a second was in the ancestry of *I. splendor-sylvae*. The third reticulation was in the ancestor of the blue clade. The third sample (Figure 4.6c) had only two reticulations, one in the ancestry of a single *I. triloba* individual and the second in the ancestor of the lineage containing the red, green, and yellow clades, as seen in Chapter 3. The fourth network (Figure 4.6d) also showed only two reticulations. The first was in the ancestry of the same *I. triloba* individual identified in the third

iteration, and the second reticulation was in the ancestor of the blue clade. Finally, the fifth iteration (Figure 4.6e) showed only one reticulation in the ancestor of the lineage containing the red, green and yellow clades. In total, three of the five randomized PhyloNet analyses implicated an ancient reticulation in the lineage leading to the last common ancestor of the red, green, and yellow clades.

### *Phylogenetic PCA*

A phylogenetic PCA (pPCA) (Revell 2009) was used to collapse variation in root traits to generate univariate traits representing variation in collinear traits. Trait measurements are displayed in Figure 4.7. Measurements of starch concentration had high standard errors, so pPCA was performed twice, both with and without starch concentration. When only including root width measurements, five principal component axes were recovered (Table 4.7). The first two principal components explained >95% of the variation in root width measurements (Figure 4.8). The first PC axis explained >85% of the variation, and primarily incorporated variation in taproot width measurements. The second PC axis explained approximately 8% of the variance in the dataset, and incorporated further variance in taproot width measurements.

When starch concentration was included in the pPCA, seven PC axes were recovered (Table 4.8). The first two PC axes explained >97% of the total variance in the dataset. The first PC axis explained >92% of the variance, and incorporated root width measurements and thin lateral root starch concentration. The second PC axis explained approximately 5% of the total variance, and 68% of the variance in taproot starch concentration was explained by this component.

### *Phylogenetic signal*

Genome size, taproot width, and width of the thickest lateral root showed significant phylogenetic signal measured with both Blomberg's K and Pagel's lambda (Table 4.9).

Furthermore, taproot width where the fourth and tenth lateral roots emerged as well as taproot starch concentration exhibited a significant lambda value but not a significant K value.

Furthermore, the phylogenetic PC2 axis where starch concentration was not included in the phylogenetic PCA exhibited a significant lambda and K value (Table 4.9).

### *Correlated evolution of genome size and root traits*

Linear regression of log-transformed genome size and the first PC axis from both phylogenetic principal components analyses resulted in a significant positive correlation ( $p < 0.05$ ) (Table 4.10, Figure 4.9). However, the correlation was very weak, with  $R^2$  values of 0.0851 and 0.185 when starch included and removed, respectively (Table 4.10, Figure 4.9).

Linear regression of log-transformed genome size and the second PC axes was non-significant when starch concentration was included and removed from the pPCA.

## **Discussion**

### *Phylogenetic relationships in the Batatas complex*

Phylogenetic analyses of the Batatas complex have, in the past, shown many conflicting results. Similar to previous results, we find that *I. batatas* (with the exception of a single *I. batatas* individual) and *I. trifida* are close relatives. However, we also find that *I. ramosissima* is a close relative of *I. batatas* and *I. trifida*, in contrast to previous results. (Austin 1988b; Jarret et al. 1992; Rajapakse et al. 2004). Similar to previous results, we do find that *I. cordatotriloba*, *I. lacunosa*, and *I. tenuissima* are close relatives (Austin 1988b; Jarret et al. 1992). Additionally, we find that the newly described species, *I. austinii*, which was described based on populations

collected in North and South Carolina (Duncan and Rausher 2013), is a close relative of two members of the Batatas complex whose ranges are limited to South America, *I. cynanchifolia* and *I. grandifolia* (Khoury et al. 2015). A formal species description of *Ipomoea austinii* has not yet been made but is in progress (*pers. comm.*, Mark Rausher).

In contrast to previous studies, we found that eight of the nine species where multiple individuals were sampled were paraphyletic. Only *Ipomoea austinii* was found to be monophyletic. Surprisingly, *I. batatas* was also paraphyletic. Five of the six *I. batatas* individuals resided in a single clade; however, a single diploid *I. batatas* individual from Tamaulipas, Mexico was found to be sister to most of the Batatas complex, suggesting that the diagnostic features of *I. batatas* have been independently evolved in this species complex.

#### *Reticulate evolution*

One major goal of this research was to characterize the extent and timing of hybridization in the Batatas complex. Therefore, we applied two methods that estimate hybridization in the face of incomplete lineage sorting. However, these methods infer hybridization events in different ways. First, HyDe uses phylogenetic invariants (Felsenstein and Cavender 1987) to test for asymmetrical site patterns in sets of four taxa, three ingroup and one outgroup. Similar to the logic of the ABBA-BABA test, also known as Patterson's D statistic (Patterson et al. 2010; Durand et al. 2011), equal frequencies of a particular site pattern indicate incomplete lineage sorting, while asymmetrical site patterns are suggestive of hybridization. HyDe does this by analyzing sites in an alignment. One primary assumption is that sites are independent of one another, but the HyDe analysis has been tested on multilocus gene alignments and is thought to be robust to violation of this assumption (Kubatko and Chifman 2015). In contrast, PhyloNet estimates the most likely phylogenetic network given a set of gene trees. This method takes a

phylogenetic approach to estimate reticulation events given a model of incomplete lineage sorting. In PhyloNet, the primary assumption is that gene trees are estimated without error; however, gene tree estimation error is a known issue in coalescent based phylogenetic methods (Roch and Warnow 2015).

Several previous studies have reported that many species in the Batatas complex are able to hybridize with one another (Jones and Deonier 1965; Diaz et al. 1996), and two species, *I. leucantha* and *I. grandifolia*, have been hypothesized to be hybrid taxa (Abel and Austin 1981; Austin 1988b). It is clear that barriers to gene flow are incomplete in many of the named species in the Batatas complex, but what is less apparent is whether hybridization is ongoing, ancient, or both. Our results suggest that hybridization occurred at least once in the ancestor of the lineage containing the red, yellow, and green clades. Three of the five PhyloNet analyses recover this pattern (Figure 4.6). Furthermore, the hybrids inferred from HyDe are all concentrated in the red, yellow, and green clades. Parents of the inferred hybrids were distributed across the phylogeny and from very different geographic areas, all suggestive of ancient hybridization. The PhyloNet analyses further implicate hybridization in the ancestor of the red clade (Figure 4.6a) and in the ancestry of single individuals (Figure 4.6b-d). Patterns of more recent hybridization in the red, yellow, and green clades would be obscured by ancient hybridization in the HyDe analysis, and the recent observation that there may be ongoing gene flow among sympatric populations of *I. cordatotriloba* and *I. lacunosa* (Duncan and Rausher 2013) may indicate that there has been a small degree of ongoing introgressive gene flow among some species in addition to at least one more ancient reticulation event in the ancestor of this clade.

Contrary to the results of Chapter 3, we did not infer hybrid ancestry in any *Ipomoea batatas* sample. This may be for a number of reasons. First, polyploid *I. batatas* may be the

result of an autopolyploidy event. Second, polyploid *I. batatas* may have emerged from an allopolyploidization event involving very closely related lineages. Third, *I. batatas* may have formed from an allopolyploidization event involving distantly related parents, but our analyses failed to detect hybrid ancestry. The analyses performed in Chapter 3 did show hybrid ancestry in the evolutionary history of hexaploid *I. batatas*. The HyDe analysis applies a Bonferroni correction to account for multiple comparisons, and the Bonferroni method is widely known to be a conservative approach (e.g. Moran 2003; Narum 2006). One hexaploid and one tetraploid *I. batatas* accession were identified as hybrid with a p-value close to but not below the significance threshold applied here. In the future, it would be fruitful to explore the false-negative rate using the Bonferroni correction in coordination with the authors of HyDe.

Genomic studies of both ancient and recent tetraploids have shown that genomes tend to fractionate, a process sometimes referred to as diploidization, where one of the two copies of a gene tends to be removed from the genome (Schnable et al. 2011). In many cases, fractionation occurs in a biased fashion, where genes tend to be lost from one duplicated genomic region rather than the other, and biased fractionation has been observed in many different plant groups and in tetraploids that range from 80 years to 10 million years old (Buggs et al. 2010; Schnable et al. 2011; Chen et al. 2013). It is possible that biased fractionation has also occurred in polyploid *I. batatas*. I evaluated a random sample of 50 gene trees from this study and found that in 16% of trees, an *I. batatas* individual was found to be more closely related to a member of the red, yellow or green clades than to a member of the blue clade; whereas, *I. batatas* was a member of the blue clade in 64% of gene trees. These genes were inferred to be single copy, so we expect these to be the result of genome fractionation. Because there were so few gene trees where *I. batatas* was closely related to a member of either the red, yellow, or green clades, it is

possible that biased genome fractionation occurred following polyploidization in *I. batatas*.

Future genomic studies of tetraploid and hexaploid sweetpotato should explore the possibility for genome fractionation in further detail.

#### *Correlated evolution of genome size and root traits*

Despite the observed trend that species with underground storage organs have larger genomes than those that do not store carbohydrates belowground (Grime and Mowforth 1985; Veselý et al. 2012), we did not recover this trend in the Batatas complex. Although we found a significant positive correlation between genome size and root traits, the linear regression accounts for little of the variation seen in genome size and root traits (Figure 4.9). This suggests that increases in genome size due to polyploidy or factors such as retrotransposon proliferation have very little influence on the evolution of storage roots. Instead, species with larger roots with higher starch contents were often found to have smaller genome sizes than the polyploid individuals included in this experiment.

The results instead point to the evolutionarily labile nature of storage root formation. Many traits with ecological and adaptive significance, such as flower color and seed characteristics, are evolutionarily labile in morning glories (Manos et al. 2001; Eserman et al. 2014). In fact, a previous study noted a similar pattern with respect to storage root formation across the morning glory tribe Ipomoeae (Eserman et al. 2014). Storage root formation has considerable ecological importance. Underground starch storage is a trait closely tied to life history such that perennial species store starch underground and, when vegetation is lost to environmental conditions such as freezing or fire, shoots resprout from underground carbohydrate stores (De Souza and Viera Da Silva 1987; Pate et al. 1990; Bell and Ojeda 1999).

### *Reticulate evolution and the evolution of storage roots*

Hybridization analyses performed here using both site-based and gene tree based analyses show that ancient introgression played a significant role in the evolution of the Batatas complex. The genomic and thus phenotypic consequences of hybridization has long been, and still remains, an active area of research (Grant 1981; Arnold et al. 1990; Abbott et al. 2016) and is still not well understood in many plant groups. Recent studies have shown that more distantly related species share morphological similarity due to introgression of loci controlling a particular phenotype (The Heliconius Genome Consortium et al. 2012). Even less clear is how to appropriately model phenotypic evolution on a phylogenetic network (Hahn and Nakhleh 2015). Future advances in phylogenetic network estimation as well as development of models to characterize the evolution of traits on phylogenetic networks will be imperative be able to more accurately describe the evolutionary history of so many groups of species whose history has been shaped by introgression.

### *Conclusions*

In conclusion, we find understanding the evolutionary history of the Batatas complex is confounded by many complicating factors, such as hybridization and polyploidy. It is clear from the results presented here that (1) ancient reticulation may have contributed to the early diversification of the red, yellow, and green clades; (2) taxonomic descriptions of species do not diagnose monophyletic species, suggesting parallel evolution or the influence of gene flow on traits currently used to describe species; (3) root traits which are related to starch storage do exhibit some phylogenetic inertia but generally are evolving independent of genome size increases due to polyploidy; and (4) *Ipomoea batatas* does not show strong evidence of hybrid origin in the single copy genes used for this study, despite being polyploidy. The last point may

be the result of autopolyploidy, allopolyploidy involving very closely related lineages, or allopolyploidy involving distantly related parents with subsequent biased genome fractionation. Future taxonomic studies of the Batatas complex will require careful consideration of hybridization. Furthermore, more fully characterizing the relationship between genome size and storage root formation will require more careful description of the traits required to form storage roots. Specifically, a time course study examining root shape change over time as well as measuring the temporal variation in root starch deposition will be necessary.

## References

- Abbott R.J., Barton N.H., Good J.M. 2016. Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* 25:2325–2332.
- Abel W.E., Austin D.F. 1981. Introgressive hybridization between *Ipomoea trichocarpa* and *Ipomoea lacunosa* (Convolvulaceae). *Bull. Torrey Bot. Club.* 108:231–239.
- Arnold M.L., Bennett B.D., Zimmer E.A. 1990. Natural Hybridization between *Iris fulva* and *Iris hexagona*: Pattern of Ribosomal DNA Variation. *Evolution.* 44:1512–1521.
- Artschwager E. 1924. On the anatomy of the sweet potato with notes on internal breakdown. *J. Agric. Res.* 27:157–166.
- Austin D.F. 1988. The taxonomy evolution and genetic diversity of sweetpotatoes and related wild species. In: Gregory P., editor. *Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources: Report of the 1st sweet potato planning conference 1987.* Lima, Peru: CIP. p. 27–59.
- Bell T.L., Ojeda F. 1999. Underground starch storage in *Erica* species of the Cape Floristic Region - Differences between seeders and resprouters. *New Phytol.* 144:143–152.
- Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution.* 57:717–745.
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120.
- Bucksch A., Burrige J., York L.M., Das A., Nord E., Weitz J.S., Lynch J.P. 2014. Image-Based High-Throughput Field Phenotyping of Crop Roots. *Plant Physiol.* 166:470–486.
- Buggs R.J.A., Chamala S., Wu W., Gao L., May G.D., Schnable P.S., Soltis D.E., Soltis P.S., Barbazuk W.B. 2010. Characterization of duplicate gene evolution in the recent natural

- allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol. Ecol.* 19:132–146.
- Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10:1.
- Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17:540–552.
- Chen X., Li X., Zhang B., Xu J., Wu Z., Wang B., Li H., Younas M., Huang L., Luo Y., Wu J., Hu S., Liu K. 2013. Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in allotetraploid *Brassica napus*. *BMC Genomics.* 14:346.
- Chifman J., Kubatko L. 2014. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics.* 30:3317–3324.
- Diaz J., Schmiediche P., Austin D.F. 1996. Polygon of crossability between eleven species of *Ipomoea* section *Batatas* (Convolvulaceae). *Euphytica.* 88:189–200.
- Duncan T.M., Rausher M.D. 2013. Morphological and genetic differentiation and reproductive isolation among closely related taxa in the *Ipomoea* series *Batatas*. *Am. J. Bot.* 100:1–11.
- Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.
- Edwards G.A., Endrizzi J.E. 1975. Cell size, nuclear size, and DNA content relationships in *Gossypium*. *Can. J. Genet. Cytol.* 17:181–186.
- Eserman L.A., Jarret R.L., Leebens-Mack J.H. *in prep.* Parallel evolution of storage roots in morning glories.
- Eserman L.A., Tiley G.P., Jarret R.L., Leebens-Mack J.H., Miller R.E. 2014. Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* 101:92–103.
- Felsenstein J., Cavender J.A. 1987. Invariants of phylogenies in a simple case with discrete states. *J. Classif.* 4:57–71.
- Grant V. 1981. *Plant Speciation*. New York: Columbia University Press.
- Grime J.P., Mowforth M.A. 1985. Variation in genome size – an ecological interpretation. *Nature.* 299:151–153.
- Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Philip D., Bowden J., Couger M.B., Eccles D., Li B., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C., Henschel R., LeDuc R.D., Friedman N., Regev A.

2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* 8:1–43.
- Hahn M.W., Nakhleh L. 2015. Irrational exuberance for resolved species trees. *Evolution*:7–17.
- Hansen J., Moller I. 1975. Percolation of starch and soluble carbohydrates from plant tissue for quantitative determination with anthrone. *Anal. Biochem.* 68:87–94.
- Heyduk K., Trapnell D.W., Barrett C.F., Leebens-Mack J.I.M. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linn. Soc.* 117:106–120.
- Jarret R.L., Gawel N., Whittemore A.T. 1992. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* L.] Lam.]. *J. Am. Soc. Hortic. Sci.* 117:633–637.
- Jones A., Deonier M.T. 1965. Interspecific Crosses Among *Ipomoea lacunosa*, *I. ramosa*, *I. trichocarpa*, and *I. triloba*. *Bot. Gaz.* 126:226–232.
- Khoury C.K., Heider B., Castaneda-Alvarez N.P., Achicanoy H.A., Sosa C.C., Miller R.E., Scotland R.W., Wood J.R.I., Rossel G., Eserman L.A., Jarret R.L., Yencho G.C., Bernau V., Juarez H., Sotelo S., de Haan S., Struik P.C. 2015. Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series Batatas]. *Front. Plant Sci.*
- Knight C.A., Ackerly D.D. 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecol. Lett.* 5:66–76.
- Knight C.A., Beaulieu J.M. 2008. Genome size scaling through phenotype space. *Ann. Bot.* 101:759–66.
- Kubatko L., Chifman J. 2015. An Invariants-based Method for Efficient Identification of Hybrid Species From Large-scale Genomic Data. *bioRxiv*:34348.
- Lowe S.B., Wilson L.A. 1974a. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 2. Interrelationships between tuber shape and yield. *Ann. Bot.* 38:319–326.
- Lowe S.B., Wilson L.A. 1974b. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 1. Tuber initiation, tuber growth and partition of assimilate. *Ann. Bot.* 38:307–317.
- Loytynoja A., Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci.* 102:10557–10562.
- Löytynoja A., Goldman N. 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science* 320:1632–1635.

- Manos P.S., Miller R.E., Wilkin P. 2001. Phylogenetic Analysis of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* Suggests a Generalized Model of Morphological Evolution in Morning Glories. *Syst. Bot.* 26:585–602.
- McDonald J.A. 1994. Convolvulaceae II. In: Sosa V., editor. *Flora de Veracruz*. p. 1–133.
- McDonald J.A., Hansen D.R., McDill J.R., Simpson B.B. 2011. A Phylogenetic Assessment of Breeding Systems and Floral Morphology of North American *Ipomoea* (Convolvulaceae). *J. Bot. Res. Inst. Texas.* 5:159–177.
- Mirarab S., Warnow T. 2015. ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics.* 31:i44–i52.
- Mooers A.O., Schluter D. 1999. Reconstructing Ancestor States with Maximum Likelihood: Support for One- and Two-Rate Models. *Syst. Biol.* 48:623–633.
- Moran M.D. 2003. Arguments for Rejecting the Sequential Bonferroni in Ecological Studies. *Oikos.* 100:403–405.
- Münkemüller T., Lavergne S., Bzeznik B., Dray S., Jombart T., Schiffrers K., Thuiller W. 2012. How to measure and test phylogenetic signal. *Methods Ecol. Evol.* 3:743–756.
- Narum S.R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* 7:783–787.
- Noh S.A., Lee H.-S., Kim Y.-S., Paek K.-H., Shin J.S., Bae J.M. 2013. Down-regulation of the IbEXP1 gene enhanced storage root development in sweetpotato. *J. Exp. Bot.* 64:129–142.
- Oren R., Schulze E.-D., Werk K.S., Meyer J., Schneider B.U., Heilmeyer H. 1988. Performance of two *Picea abies* (L.) Karst. stands at different stages of decline. I. Carbon relations and stand growth. *Oecologia.* 75:25–37.
- Ozias-Akins P., Jarret R.L. 1994. Nuclear DNA Content and Ploidy Levels in the Genus *Ipomoea*. *J. Soc. Hortic. Sci.* 119:110–115.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. London, Biol. Sci.* 255:37–45.
- Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature.* 401:877–884.
- Pate J.S., Friend R.H., Bowen B.J., Hansen A., Kuo J. 1990. Seedling growth and storage characteristics of seeder and resprouter species of Mediterranean-type ecosystems of SW Australia. *Ann. Bot.* 65:585–601.
- Patterson N., Petersen D.C., van der Ross R.E., Sudoyo H., Glashoff R.H., Marzuki S., Reich D., Hayes V.M. 2010. Genetic structure of a unique admixed population: Implications for medical research. *Hum. Mol. Genet.* 19:411–419.

- Rajapakse S., Nilmalgoda S.D., Molnar M., Ballard R.E., Austin D.F., Bohac J.R. 2004. Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear beta-amylase gene sequences. *Mol. Phylogenet. Evol.* 30:623–32.
- Reddy U.K., Bates G.T., Ryan-Bohac J., Nimmakayala P. 2007. The Sweetpotato. In: Cole K., editor. *Genome Mapping and Molecular Breeding in Plants*. Berlin: Springer-Verlag. p. 237–247.
- Revell L.J. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution*. 63:3258–3268.
- Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.
- Roch S., Warnow T. 2015. On the Robustness to Gene Tree Estimation Error (or lack thereof) of Coalescent-Based Species Tree Methods. *Syst. Biol.* 64:663–676.
- Rohlf J. 2010. tpsDIG2. Available from <http://life.bio.sunysb.edu/morph/>.
- Roullier C., Duputié A., Wennekes P., Benoit L., Fernández Bringas V.M., Rossel G., Tay D., McKey D., Lebot V. 2013. Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS One*. 8:e62707.
- Schnable J.C., Springer N.M., Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci.* 108:4069–4074.
- Shapiro S.S., Francia R.S. 1972. An approximate of variance test analysis for normality. *J. Am. Stat. Assoc.* 67:215–216.
- De Souza J.G., Viera Da Silva J. 1987. Partitioning of carbohydrates in annual and perennial cotton (*Gossypium hirsutum* L.). *J. Exp. Bot.* 38:1211–1218.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Talavera G., Castresana J. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* 56:564–577.
- Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*. 9:322.
- The Heliconius Genome Consortium, Dasmahapatra K.K., Walters J.R., Briscoe A.D., Davey J.W., Whibley A., Nadeau N.J., Zimin A. V., Hughes D.S.T., Ferguson L.C., Martin S.H., Salazar C., Lewis J.J., Adler S., Ahn S.-J., Baker D.A., Baxter S.W., Chamberlain N.L., Chauhan R., Counterman B.A., Dalmay T., Gilbert L.E., Gordon K., Heckel D.G., Hines

- H.M., Hoff K.J., Holland P.W.H., Jacquin-Joly E., Jiggins F.M., Jones R.T., Kapan D.D., Kersey P., Lamas G., Lawson D., Mapleson D., Maroja L.S., Martin A., Moxon S., Palmer W.J., Papa R., Papanicolaou A., Pauchet Y., Ray D.A., Rosser N., Salzberg S.L., Supple M.A., Surridge A., Tenger-Trolander A., Vogel H., Wilkinson P.A., Wilson D., Yorke J.A., Yuan F., Balmuth A.L., Eland C., Gharbi K., Thomson M., Gibbs R.A., Han Y., Jayaseelan J.C., Kovar C., Mathew T., Muzny D.M., Onger F., Pu L.-L., Qu J., Thornton R.L., Worley K.C., Wu Y.-Q., Linares M., Blaxter M.L., Ffrench-Constant R.H., Joron M., Kronforst M.R., Mullen S.P., Reed R.D., Scherer S.E., Richards S., Mallet J., Owen McMillan W., Jiggins C.D. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 487:94–98.
- Veselý P., Bureš P., Šmarda P., Pavlíček T. 2012. Genome size and DNA base composition of geophytes: The mirror of phenology and ecology? *Ann. Bot.* 109:65–75.
- Wilson L.A., Lowe S.B. 1973. The anatomy of the root system in West Indian sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. *Ann. Bot.* 37:633–643.
- Yu Y., Dong J., Liu K.J., Nakhleh L. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proc. Natl. Acad. Sci.* 111.
- Yu Y., Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*. 16:S10.
- Zeeman S.C., Northrop F., Smith A.M., ap Rees T. 1998. A starch-accumulating mutant of *Arabidopsis thaliana* deficient in a starch-hydrolyzing enzyme. *Plant J.* 15:357–365.

Table 4.1. Accession information for the individuals sequenced in this study.

Species	Accession	Locality	Seed Source
<i>I. austinii</i>	U14	USA-North Carolina	T. Duncan
<i>I. austinii</i>	U1a7	USA-North Carolina	T. Duncan
<i>I. batatas</i>	PI518474	Mexico-Veracruz	USDA GRIN
<i>I. batatas</i>	PI561261	Ecuador-Loja	USDA GRIN
<i>I. batatas</i>	REM356	Mexico-Jalisco	R. E. Miller
<i>I. batatas</i> cv. <b>Beauregard</b>	PI566613	cultivated	USDA GRIN
<i>I. batatas</i> cv. <b>Jewel</b>	PI531122	cultivated	USDA GRIN
<i>I. cordatotriloba</i>	PI518494	Mexico-Tabasco	USDA GRIN
<i>I. cordatotriloba</i>	PI518495	Mexico-Tabasco	USDA GRIN
<i>I. cordatotriloba</i>	REM317	USA-Louisiana	R. E. Miller
<i>I. cordatotriloba</i>	REM857	USA-Mississippi	R. E. Miller
<i>I. cordatotriloba</i>	REM860	USA-Alabama	R. E. Miller
<i>I. cordatotriloba</i>	REM861	USA-Alabama	R. E. Miller
<i>I. cynanchifolia</i>	PI549093	Brazil	USDA GRIN
<i>I. grandifolia</i>	PI561549	Peru-Lima	USDA GRIN
<i>I. grandifolia</i>	PI561550	Peru-Lima	USDA GRIN
<i>I. lacunosa</i>	PI634785	USA-South Carolina	USDA GRIN
<i>I. lacunosa</i>	Q1149	Japan-Aichi	E. Nitasaka
<i>I. lacunosa</i>	REM262	USA-North Carolina	R. E. Miller
<i>I. lacunosa</i>	REM310	USA-North Carolina	K. Iwao
<i>I. leucantha</i>	PI518481	Mexico-Tabasco	USDA GRIN
<i>I. leucantha</i>	PI536036	Mexico-Veracruz	USDA GRIN
<i>I. leucantha</i>	PI540733	Colombia-Cesar	USDA GRIN
<i>I. leucantha</i>	PI540735	Colombia-Cesar	USDA GRIN
<i>I. leucantha</i>	R8P	USA-South Carolina	T. Duncan
<i>I. ramosissima</i>	PI540711	Colombia-Cesar	USDA GRIN
<i>I. sepacuitensis</i>	LAE69	cultivated	B&T World Seeds
<i>I. trifida</i>	E/PAu23	Panama-Gamboa	E. Eich
<i>I. trifida</i>	E/Pau27	Panama-Isla Taboga	E. Eich
<i>I. trifida</i>	PI543830	Costa Rica-Playa Ocotal	USDA GRIN
<i>I. trifida</i>	PI561543	Venezuela	USDA GRIN
<i>I. trifida</i>	PI561547	Guatemala	USDA GRIN
<i>I. trifida</i>	REM364	Mexico-Jalisco	R. E. Miller
<i>I. trifida</i>	REM416	Mexico-Nayarit	R. E. Miller
<i>I. trifida</i>	REM450	Mexico-Oaxaca	R. E. Miller
<i>I. trifida</i>	REM753	Costa Rica-Guanacaste	R. E. Miller
<i>I. triloba</i>	PI530997	Dominican Republic-San Pedros de Macoris	USDA GRIN
<i>I. triloba</i>	PI530998	Dominican Republic-Valverde	USDA GRIN
<i>I. triloba</i>	PI536040	Mexico-Tabasco	USDA GRIN
<i>I. triloba</i>	PI536041	Mexico-Tabasco	USDA GRIN
<i>I. triloba</i>	PI536042	Mexico-Campeche	USDA GRIN
<i>I. triloba</i>	PI536044	Mexico-Chiapas	USDA GRIN
<i>I. triloba</i>	PI540710	Colombia-Cesar	USDA GRIN
<i>I. triloba</i>	PI540731	Colombia-Cesar	USDA GRIN

<i>I. triloba</i>	PI561554	Australia	USDA GRIN
<i>I. triloba</i>	PI618965	Mexico-Michoacan	USDA GRIN
<i>I. triloba</i>	PI634795	USA-South Carolina	USDA GRIN
<i>I. triloba</i>	Q1112	Japan-Saitama	E. Nitasaka
<i>I. triloba</i>	REM355	Mexico-Jalisco	R. E. Miller
<i>I. unknown</i>	LAE87	USA-Louisiana	L. A. Eserman
<i>I. unknown</i>	LAE93	USA-Louisiana	B. D. Gibbens
<i>I. unknown</i>	LAE-FL19	USA-Florida	L. A. Eserman, R. E. Miller
<i>I. unknown</i>	MTC100	Mexico-Morelos	M. T. Clegg
<i>I. unknown</i>	MTC106	Mexico-Yucatan	M. T. Clegg
<i>I. unknown</i>	MTC165	Mexico-Morelos	M. T. Clegg
<i>I. unknown</i>	MTC169	Mexico-Morelos	M. T. Clegg
<i>I. unknown</i>	MTC185	Mexico-Morelos	M. T. Clegg
<i>I. unknown</i>	MTC37	Mexico-Guererro	M. T. Clegg
<i>I. unknown</i>	PI536039	Mexico-Tabasco	USDA GRIN
<i>I. unknown</i>	REM373	Mexico-Jalisco	R. E. Miller
<i>I. unknown</i>	REM453	Mexico-Oaxaca	R. E. Miller
<i>I. unknown</i>	REM807	USA-Louisiana	R. E. Miller
<i>I. unknown</i>	REM811	USA-Louisiana	R. E. Miller
<i>I. unknown</i>	REM862	USA-Alabama	R. E. Miller

Table 4.2. Genome size measurements for accessions used in this study.

Species	Accession	Genome size (pg/2C)	Standard Deviation	Ploidy
<i>I. austinii</i>	U14			2x*
<i>I. austinii</i>	U1a7	0.92	0.014	2x
<i>I. batatas</i>	PI518474	1.83	0.018	4x
<i>I. batatas</i>	PI561261	1.70	0.013	4x
<i>I. batatas</i>	REM356	0.88	0.004	2x
<i>I. batatas</i> cv. <b>Beauregard</b>	PI566613	2.63	0.021	6x
<i>I. batatas</i> cv. <b>Jewel</b>	PI531122	2.56	0.014	6x
<i>I. cordatotriloba</i>	PI518494	‡		
<i>I. cordatotriloba</i>	PI518495	0.93	0.013	2x
<i>I. cordatotriloba</i>	REM317	0.94	0.023	2x
<i>I. cordatotriloba</i>	REM857	1.05	0.012	2x
<i>I. cordatotriloba</i>	REM860	0.97	0.015	2x
<i>I. cordatotriloba</i>	REM861	0.90	0.019	2x
<i>I. cynanchifolia</i>	PI549093	0.83	0.013	2x
<i>I. grandifolia</i>	PI561549	0.84	0.013	2x
<i>I. grandifolia</i>	PI561550	0.87	0.001	2x
<i>I. lacunosa</i>	PI634785	0.90	0.012	2x
<i>I. lacunosa</i>	Q1149	1.03	0.028	2x
<i>I. lacunosa</i>	REM262	1.03	0.013	2x
<i>I. lacunosa</i>	REM310	1.08	0.027	2x
<i>I. leucantha</i>	PI518481	0.81	0.025	2x
<i>I. leucantha</i>	PI536036	1.20	0.025	2x
<i>I. leucantha</i>	PI540733	0.95	0.030	2x
<i>I. leucantha</i>	PI540735	0.97	0.025	2x
<i>I. leucantha</i>	R8P	0.84	0.012	2x
<i>I. ramosissima</i>	PI540711	0.87	0.038	2x
<i>I. sepacuitensis</i>	LAE69	1.72	0.015	
<i>I. trifida</i>	E/PAu23	0.97	0.020	2x
<i>I. trifida</i>	P/Pau27	‡		
<i>I. trifida</i>	PI543830	‡		
<i>I. trifida</i>	PI561543	1.02	0.016	2x
<i>I. trifida</i>	PI561547	0.94	0.027	2x
<i>I. trifida</i>	REM364	0.84	0.016	2x
<i>I. trifida</i>	REM416	0.86	0.021	2x
<i>I. trifida</i>	REM450	1.08	0.029	2x
<i>I. trifida</i>	REM753	0.98	0.015	2x
<i>I. triloba</i>	PI530997	0.99	0.029	2x
<i>I. triloba</i>	PI530998	0.95	0.016	2x
<i>I. triloba</i>	PI536040	0.94	0.017	2x
<i>I. triloba</i>	PI536041	0.90	0.036	2x

<i>I. triloba</i>	PI536042	0.89	0.033	2x
<i>I. triloba</i>	PI536044	0.93	0.044	2x
<i>I. triloba</i>	PI540710	0.92	0.033	2x
<i>I. triloba</i>	PI540731	1.02	0.047	2x
<i>I. triloba</i>	PI561554	0.97	0.010	2x
<i>I. triloba</i>	PI618965	0.90	0.026	2x
<i>I. triloba</i>	PI634795	0.91	0.017	2x
<i>I. triloba</i>	Q1112	1.17	0.014	2x
<i>I. triloba</i>	REM355	0.90	0.021	2x
<i>I. unknown</i>	LAE87	0.97	0.013	2x
<i>I. unknown</i>	LAE93	0.93	0.008	2x
<i>I. unknown</i>	LAE-FL19	†		
<i>I. unknown</i>	MTC100	0.98	0.024	2x
<i>I. unknown</i>	MTC106	1.20	0.012	2x
<i>I. unknown</i>	MTC165	1.04	0.014	2x
<i>I. unknown</i>	MTC169	0.91	0.028	2x
<i>I. unknown</i>	MTC185	0.94	0.027	2x
<i>I. unknown</i>	MTC37	0.94	0.031	2x
<i>I. unknown</i>	PI536039	0.87	0.024	2x
<i>I. unknown</i>	REM373	0.89	0.015	2x
<i>I. unknown</i>	REM453	1.04	0.044	2x
<i>I. unknown</i>	REM807	0.95	0.021	2x
<i>I. unknown</i>	REM811	0.97	0.021	2x
<i>I. unknown</i>	REM862	0.93	0.019	2x

\* Accession determined to be diploid in Duncan & Rausher (2013).

† DNA isolated from dried, field collected leaf material, no seeds were available.

‡ Plant died before genome size could be measured and was removed from further analysis.

Table 4.3. Between experiment controls for flow cytometric measurement of genome size. Shown are genome size measurements and inferred ploidy level.

Species	Accession	Ozias-Akins, Jarrett (1994)		Present Experiment	
		Genome size (pg/2C)	Inferred ploidy	Genome size (pg/2C)	Inferred ploidy
<i>I. cordatotriloba</i>	PI518495	3.3	4x	0.93	2x
<i>I. cynanchifolia</i>	PI549093	1.7	2x	0.83	2x
<i>I. leucantha</i>	PI536036	1.6	2x	1.20	2x
<i>I. tabascana</i>	PI518479	2.6	4x	1.96	4x

Table 4.4. Across year controls for flow cytometric measurement of genome size. Genome size is measured in pg/2C, and each measurement includes standard error of three replicate measurements of the same leaf. *Ipomoea lacunosa* (REM844) and *I. trifida* (REM774) were initially measured for Chapter 4 and are included here as controls.

<b>Species</b>	<b>Accession</b>	<b>2015 Genome Size ± SE (pg/2C)</b>	<b>2016 Genome Size ± SE (pg/2C)</b>
<i>I. lacunosa</i>	Q1149	1.03 ±0.028	0.94 ±0.021
<i>I. lacunosa</i>	REM844	1.10 ±0.013	0.90 ±0.011
<i>I. trifida</i>	REM774	1.04 ±0.021	0.94 ±0.060
<i>I. trifida</i>	PI561547	0.94 ±0.027	0.93 ±0.014

Table 4.5. Sequencing results for all accessions sequenced in this project.

Species	Accession	Raw Reads	Filtered reads	Mean Gene Length	Mean Exon Length	Mean Intron Length	Mean Exon Cov	Mean Intron Cov
<i>I. austinii</i>	U14	24,428,658	24,428,429	1340.93	500.32	840.60	1473.22	447.14
<i>I. austinii</i>	Ula7	7,249,352	7,249,314	1604.80	587.35	1017.45	455.25	171.29
<i>I. batatas</i>	PI518474	17,537,158	17,537,046	885.43	316.70	568.73	771.52	318.65
<i>I. batatas</i>	PI561261	14,369,786	14,369,681	975.49	379.66	595.83	958.29	250.04
<i>I. batatas*</i>	REM356	3,528	3,528	186.27	170.73	13.15	2.24	2.27
<i>I. batatas</i> cv. <b>Beauregard</b>	PI566613	22,707,798	22,707,643	565.93	196.03	366.70	757.84	331.59
<i>I. batatas</i> cv. <b>Jewel</b>	PI531122	2,655,244	2,655,233	701.49	267.20	430.96	94.45	45.43
<i>I. cordatotriloba</i> ‡	PI518494	13,279,998	13,279,908	2.33	1353.67	544.43	805.97	964.13
<i>I. cordatotriloba</i>	PI518495	8,941,456	8,941,411	1514.41	580.89	929.13	585.88	202.31
<i>I. cordatotriloba</i>	REM317	15,550,018	15,549,945	1431.05	486.11	944.94	760.37	365.05
<i>I. cordatotriloba</i>	REM857	6,541,498	6,541,451	1560.62	654.97	901.96	570.65	135.64
<i>I. cordatotriloba</i>	REM860	12,536,040	12,535,972	1509.70	581.34	928.36	825.05	262.02
<i>I. cordatotriloba</i>	REM861	26,277,572	26,277,420	1341.53	469.20	868.28	1375.28	530.67
<i>I. cynanchifolia</i>	PI549093	5,112,136	5,112,106	1673.99	644.97	1029.03	337.48	115.09
<i>I. grandifolia</i>	PI561549	18,740,742	18,740,638	1361.19	437.09	919.26	793.25	425.51
<i>I. grandifolia</i>	PI561550	3,431,432	3,430,767	1478.54	708.02	770.52	287.71	63.50
<i>I. lacunosa</i>	PI634785	38,245,104	38,244,848	1199.01	441.74	753.48	2221.29	623.59
<i>I. lacunosa</i>	Q1149	10,819,764	10,819,698	1506.05	520.55	980.68	649.14	270.99
<i>I. lacunosa</i>	REM262	14,069,916	14,069,857	1248.31	414.08	834.23	664.73	348.56
<i>I. lacunosa</i>	REM310	13,952,792	13,952,715	1484.94	515.80	969.14	781.21	311.62
<i>I. leucantha</i>	PI518481	1,416,566	1,416,319	1255.15	665.61	589.54	116.00	26.71
<i>I. leucantha</i>	PI536036	27,517,218	27,517,061	1247.36	381.75	865.62	1267.30	636.99
<i>I. leucantha</i>	PI540733	19,444,570	19,444,449	1164.37	477.88	683.17	1293.80	339.87
<i>I. leucantha</i>	PI540735	10,102,518	10,102,467	1502.63	568.53	929.49	613.74	230.13
<i>I. leucantha</i>	R8P	32,655,244	32,655,040	1404.43	546.62	852.98	2343.69	656.42
<i>I. ramosissima</i>	PI540711	10,518,344	10,518,266	1628.80	601.19	1027.61	747.64	214.02
<i>I. sepacuitensis</i>	LAE69	17,802,494	17,802,405	1344.08	460.93	883.15	874.18	342.66
<i>I. trifida</i>	E/PAu23	13,284,738	13,284,645	1075.15	382.28	692.87	649.90	250.20
<i>I. trifida</i> ‡	E/PAu27	8,687,046	8,686,996	1222.57	469.19	753.38	416.48	170.07
<i>I. trifida</i> ‡	PI543830	23,025,140	23,024,998	2.054	1125.45	421.69	703.76	1409.38
<i>I. trifida</i>	PI561543	5,443,538	5,443,511	1375.85	547.84	828.01	327.63	117.11
<i>I. trifida</i>	PI561547	24,460,370	24,460,255	1128.29	377.54	745.89	1026.34	585.99
<i>I. trifida</i>	REM364	17,450,586	17,450,478	1250.58	404.66	841.79	727.19	387.16
<i>I. trifida</i>	REM416	43,059,156	43,058,929	1465.49	510.37	955.12	2370.62	965.66
<i>I. trifida</i>	REM450	7,718,694	7,718,652	1235.84	488.39	747.44	404.75	158.14
<i>I. trifida</i>	REM753	3,583,922	3,583,908	1233.45	559.99	669.60	295.76	65.68
<i>I. triloba</i>	PI530997	9,387,180	9,387,120	1471.63	539.49	932.14	601.09	210.47
<i>I. triloba</i>	PI530998	24,854,078	24,853,932	1292.11	406.02	886.09	1091.68	580.64
<i>I. triloba</i>	PI536040	11,105,930	11,105,865	1504.44	575.99	924.00	647.16	251.92
<i>I. triloba</i>	PI536041	2,438,872	2,438,856	1675.65	649.18	1026.48	171.01	65.30
<i>I. triloba</i>	PI536042	4,217,346	4,217,319	1577.96	585.75	987.89	246.53	102.62
<i>I. triloba</i>	PI536044	19,858,590	19,858,473	1419.43	499.69	919.74	1129.23	438.38

<i>I. triloba</i>	PI540710	57,480,854	57,480,569	1114.06	315.90	798.16	2252.28	1291.24
<i>I. triloba</i>	PI540731	15,555,856	15,555,767	1502.22	547.94	954.28	1035.15	337.12
<i>I. triloba</i>	PI561554	5,843,484	5,843,460	1533.32	570.41	958.82	349.44	135.83
<i>I. triloba</i>	PI618965	14,051,092	14,050,998	1563.27	585.91	973.21	1020.94	321.28
<i>I. triloba</i>	PI634795	2,035,496	2,035,490	1723.35	681.90	1041.45	140.50	53.54
<i>I. triloba</i>	Q1112	15,726,882	15,726,786	1381.72	495.87	881.04	883.97	376.36
<i>I. triloba</i>	REM355	15,922,996	15,922,901	1478.28	498.12	975.22	821.05	384.63
<i>I. unknown</i>	LAE87	21,886,410	21,886,288	1443.63	502.21	936.86	1326.34	513.26
<i>I. unknown</i>	LAE93	26,291,400	26,291,280	1347.73	487.77	855.37	1317.19	592.05
<i>I. unknown</i>	LAE- FL19	60,981,054	60,980,705	1335.35	490.43	844.92	3264.88	1196.26
<i>I. unknown</i>	MTC100	34,913,446	34,913,266	1275.81	412.59	859.07	1688.53	696.21
<i>I. unknown</i>	MTC106	29,183,114	29,182,936	1270.00	424.60	845.40	1646.33	647.43
<i>I. unknown</i>	MTC165	16,709,258	16,709,159	1412.40	491.59	916.40	949.12	397.03
<i>I. unknown</i>	MTC169	14,101,822	14,101,733	1493.38	559.96	933.42	1037.09	308.23
<i>I. unknown</i>	MTC185	10,009,324	10,009,269	1471.05	510.43	956.03	520.33	231.69
<i>I. unknown</i>	MTC37	17,949,788	17,949,685	1369.76	483.76	881.65	962.50	394.21
<i>I. unknown</i>	PI536039	7,175,654	7,175,617	1460.96	582.10	875.28	561.61	154.66
<i>I. unknown</i>	REM373	10,582,970	10,582,905	1391.27	489.81	897.18	566.70	251.33
<i>I. unknown</i>	REM453	10,119,724	10,119,672	1048.66	360.08	688.58	505.56	207.29
<i>I. unknown</i>	REM807	16,912,404	16,912,307	1449.44	507.78	937.69	980.59	373.37
<i>I. unknown</i>	REM811	7,631,764	7,631,714	1583.48	641.23	942.26	619.12	176.03
<i>I. unknown</i>	REM862	22,766,934	22,766,785	1387.20	471.54	910.94	1150.77	522.89

\* Sample removed for low sequencing depth.

‡ Plant died before genome size could be measured and was removed from further analysis.

Table 4.6. Alignment statistics for the two datasets used in this study.

<b>Taxa in dataset</b>	<b>No. Individuals</b>	<b>No. Genes</b>	<b>Mean gene length</b>	<b>Total aligned length</b>	<b>Parsimony uninformative sites</b>	<b>Parsimony informative sites</b>	<b>Mean missing data</b>
<b>Diploids</b>	71	244	1409.07	505,689	44,297	25,335	27.8%
<b>Diploids + Polyploids</b>	77	244	1350.70	490,520	43,600	25,849	28.6%

Table 4.7. Proportion of variance in root traits explained by the five phylogenetic principal component axes for the pPCA that did not include starch measurements. Width at the fourth lateral root and width of the thinnest lateral root were transformed by squaring raw data points.

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>
<b>Width of taproot</b>	0.924	0.0177	0.0584	8.74E-05	1.84E-11
<b>Width of taproot - 4th lateral</b>	0.749	0.179	0.0717	2.75E-05	7.63E-11
<b>Width of taproot - 10th lateral</b>	0.903	0.0889	0.00817	1.25E-06	3.70E-13
<b>Thick lateral root width</b>	0.0439	0.000245	0.0167	0.939	1.76E-10
<b>Thin lateral root width</b>	0.00175	0.00495	0.0179	0.000197	0.975
<b>Total variance explained by PC axis</b>	0.871	0.0808	0.0393	0.0086	0.0000066

Table 4.8. Proportion of variance in root traits explained by the five phylogenetic principal component axes for the pPCA that included starch measurements. Width at the fourth lateral root and width of the thinnest lateral root were transformed by squaring data, and both measures of starch concentration were square-root transformed.

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>	<b>PC7</b>
<b>Width of taproot</b>	0.227	0.019	0.667	0.001	0.086	0.000	1.34E-09
<b>Width of taproot - 4th lateral</b>	0.265	0.060	0.513	0.138	0.023	2.97E-06	1.97E-06
<b>Width of taproot - 10th lateral</b>	0.126	0.007	0.776	0.073	0.018	0.000	5.38E-08
<b>Thick lateral root width</b>	0.002	0.028	0.054	9.74E-05	0.020	0.896	0.000192
<b>Thin lateral root width</b>	0.037	0.012	0.023	0.021	0.003	0.007	0.897
<b>Taproot starch concentration</b>	0.318	0.681	0.002	6.09E-05	1.10E-06	1.87E-07	9.27E-10
<b>Thin lateral root starch concentration</b>	0.998	0.001	0.000	1.35E-06	2.65E-09	4.15E-10	1.23E-10
<b>Total variance explained by PC axis</b>	0.924	0.0547	0.0181	0.00186	0.00112	0.000197	3.14E-05

Table 4.9. Measures of phylogenetic signal for the traits used in this study. Phylogenetic signal was calculated using Blomberg's K and Pagel's lambda statistics. pPC1 and pPC2 are the first two axes from the phylogenetic principal components analysis. Underlined K and lambda values are those found to be significant using 1000 bootstrap randomizations. Genome size was log-transformed, width at the fourth lateral root and width of the thinnest lateral root were squared, and both measures of starch concentration were square-root transformed.

	<b>Blomberg's K</b>		<b>Pagel's lambda</b>	
	<b>K</b>	<b>P-value</b>	<b>Lambda</b>	<b>P-value</b>
<b>Genome Size</b>	<u>1.754</u>	0.002	<u>1.000</u>	2.98E-18
<b>Width of taproot</b>	<u>0.473</u>	0.036	<u>0.773</u>	0.00012
<b>Width of taproot - 4th lateral</b>	0.429	0.058	<u>0.768</u>	4.66E-05
<b>Width of taproot - 10th lateral</b>	0.244	0.196	<u>0.583</u>	0.0171
<b>Thin lateral</b>	0.155	0.596	0.000	1.000
<b>Thick lateral</b>	<u>0.436</u>	0.047	<u>0.690</u>	0.00199
<b>Thin lateral root starch concentration</b>	0.120	0.921	0.388	0.211
<b>Taproot starch concentration</b>	0.289	0.124	<u>0.787</u>	0.00309
<b>PC1 – no starch</b>	0.242	0.186	<u>0.671</u>	0.0132
<b>PC2 – no starch</b>	<u>1.489</u>	0.002	<u>0.852</u>	1.64E-11
<b>PC1 – with starch</b>	0.114	0.926	0.416	0.161
<b>PC2 – with starch</b>	0.233	0.216	0.669	0.110

Table 4.10. Results of the linear regression of phylogenetic independent contrasts of log-transformed genome size and the first two axes of the phylogenetic principal components analysis. Shown are results with and without starch concentration included in the pPCA.

	<b>Without Starch Concentration</b>			<b>With Starch Concentration</b>		
	F <sub>1,61</sub>	P-value	R <sup>2</sup>	F <sub>1,61</sub>	P-value	R <sup>2</sup>
<b>PC1</b>	13.85	0.000433	0.185	5.673	0.0204	0.0851
<b>PC2</b>	0.131	0.719	0.00215	0.390	0.535	0.00636

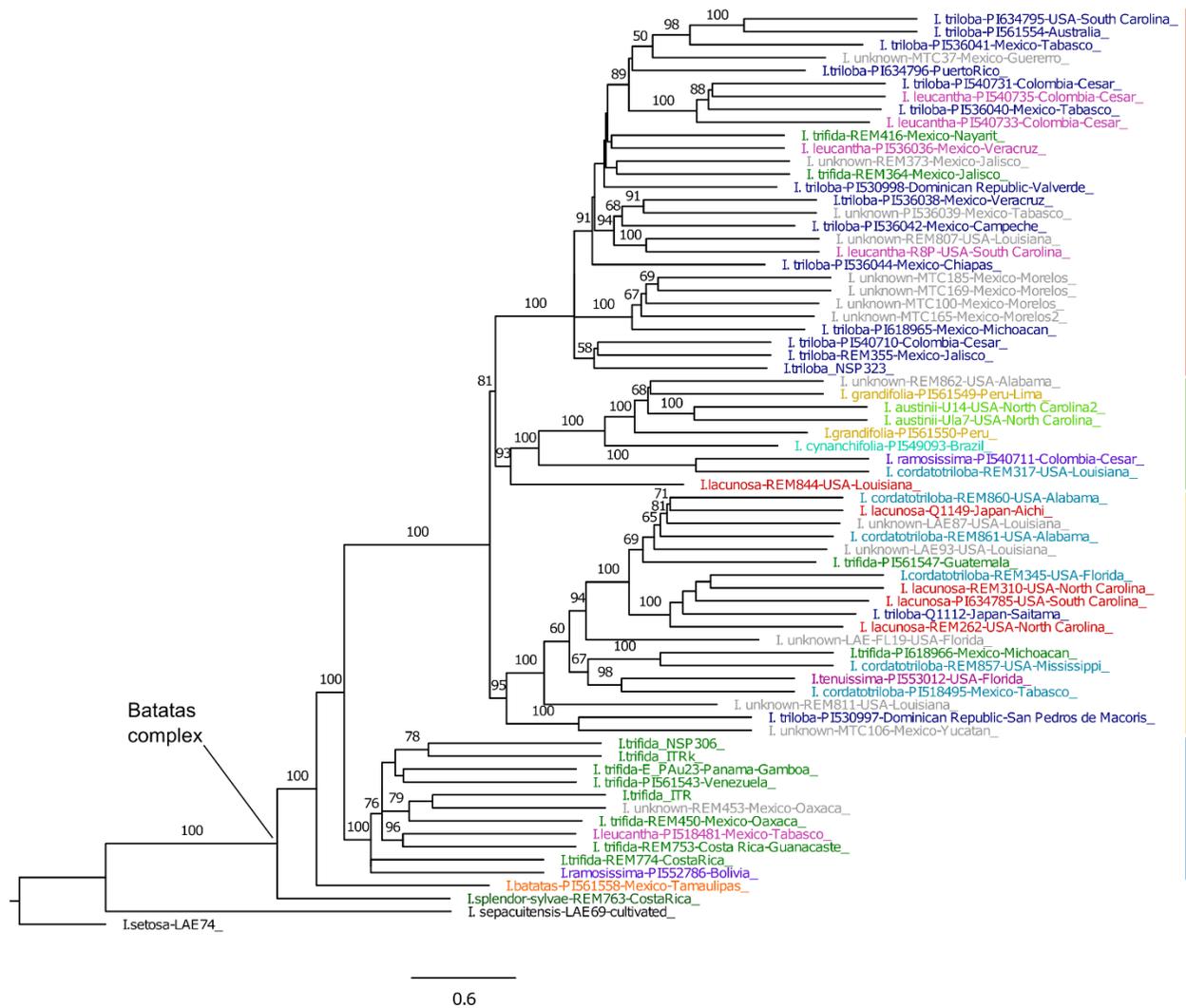


Figure 4.1. Phylogeny of the Batatas complex estimated in ASTRAL-II including only diploid individuals. Numbers behind nodes indicate bootstrap support from multilocus bootstrapping done in ASTRAL-II. Nodes without numbers had bootstrap support <50%. Each individual is colored by species designation. Grey individuals are those which exhibited intermediate traits and could not confidently be identified to species using existing taxonomic keys. Colored bars to the right of the phylogeny indicate well-supported clades recovered in the concatenation, ASTRAL-II, and SVDQuartets analyses.



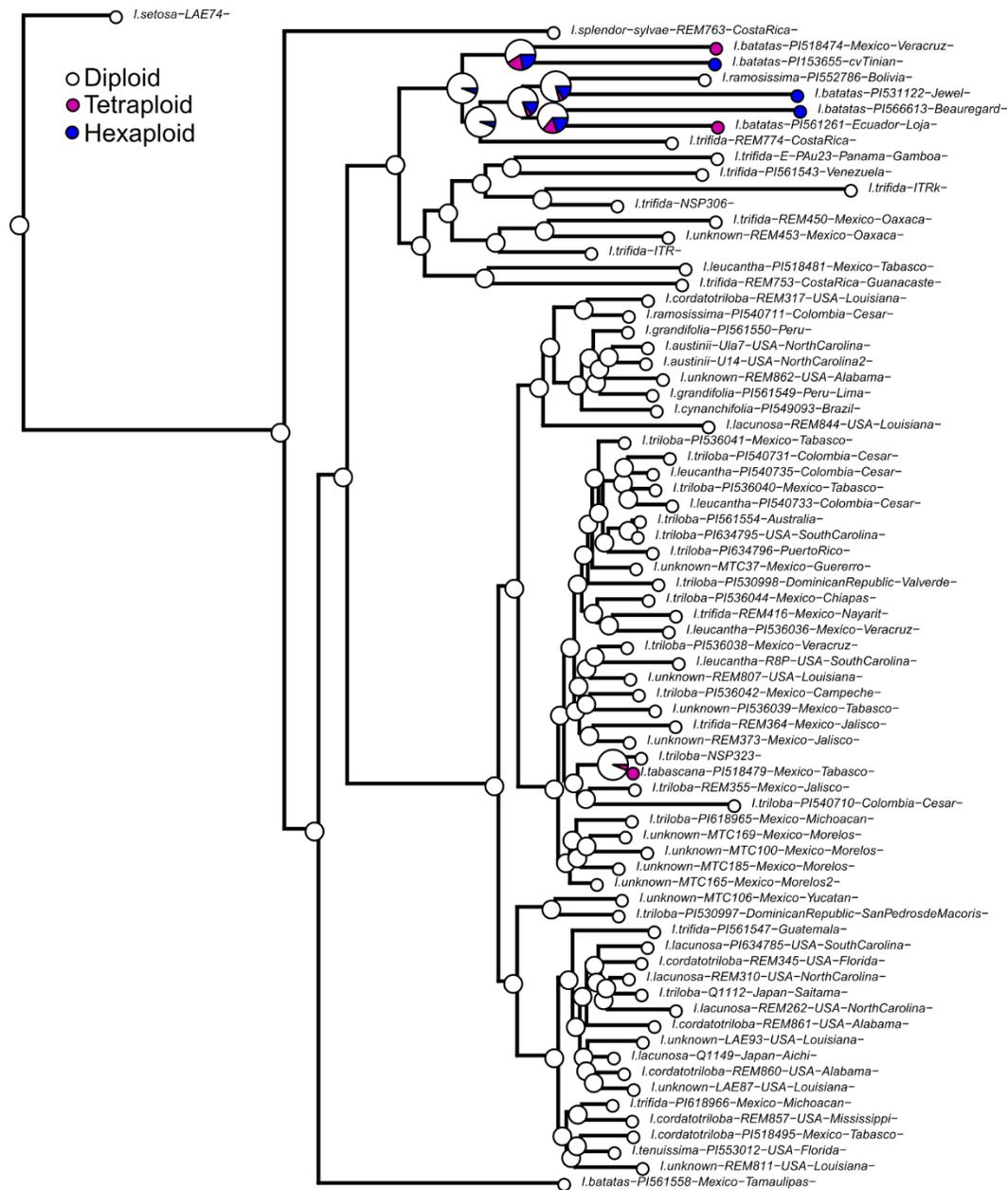


Figure 4.3. Evolutionary history of polyploidy in the Batatas complex. Open circles indicate diploid taxa, purple indicates tetraploids, and blue indicates hexaploids. Larger pie charts are on nodes with a non-zero probability of having an ancestral character state other than diploid.

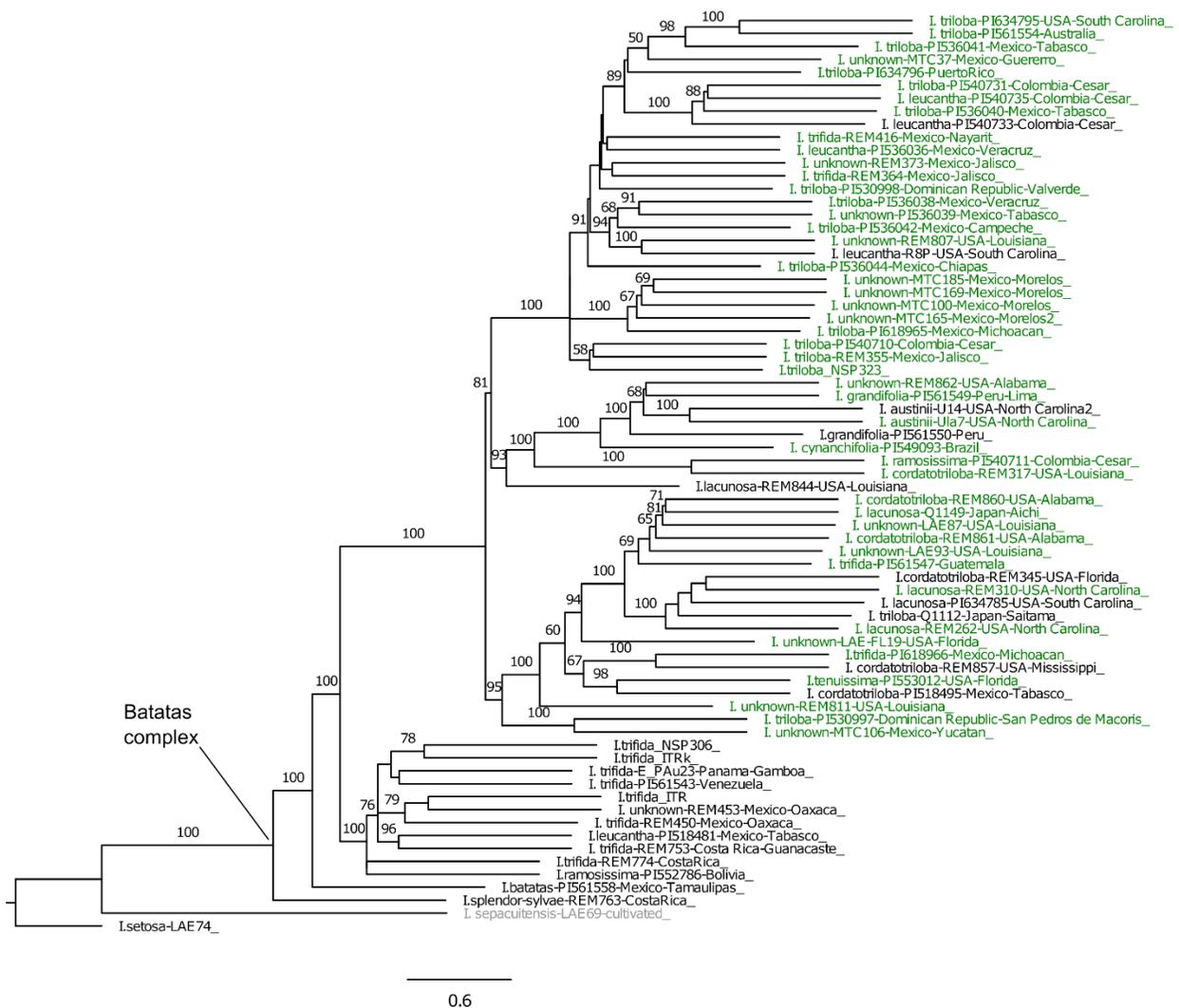


Figure 4.4. Phylogeny of the Batatas complex estimated in ASTRAL-II illustrating results from the HyDe analysis of diploid taxa. Individuals in green were inferred to be hybrids in HyDe. *Ipomoea sepacuitensis* is shown in grey because it was removed from the HyDe analysis.



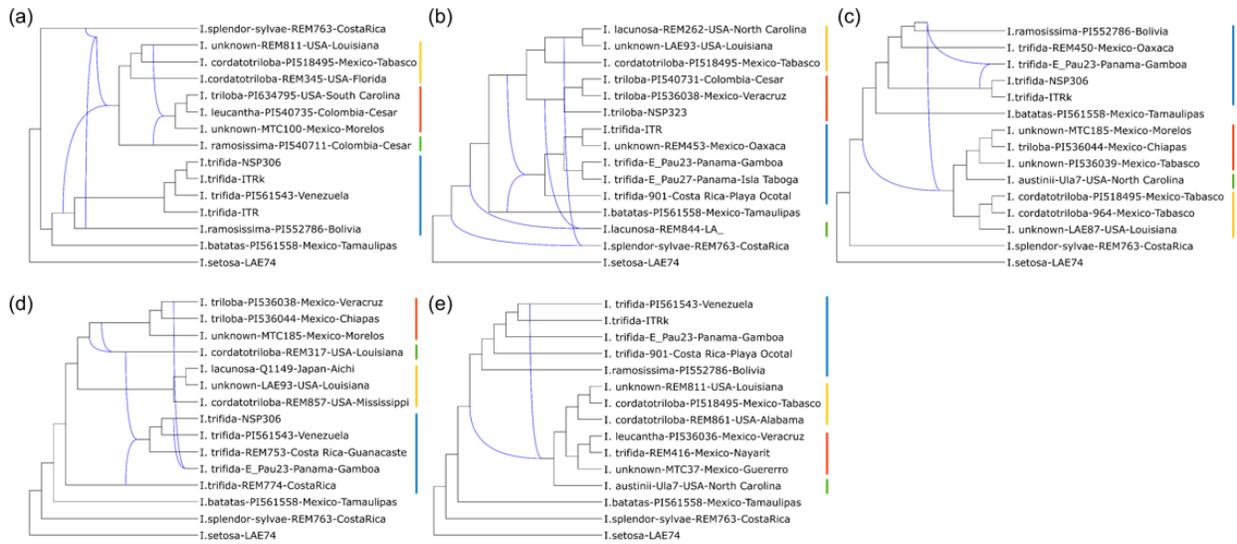


Figure 4.6. Results of the PhyloNet analysis of five random samples of diploid taxa. Each network represents the most likely network given the particular set of gene trees. Each dataset includes *I. setosa* as an outgroup, *I. splendora-sylvae* (REM763) and *I. batatas* PI561558.

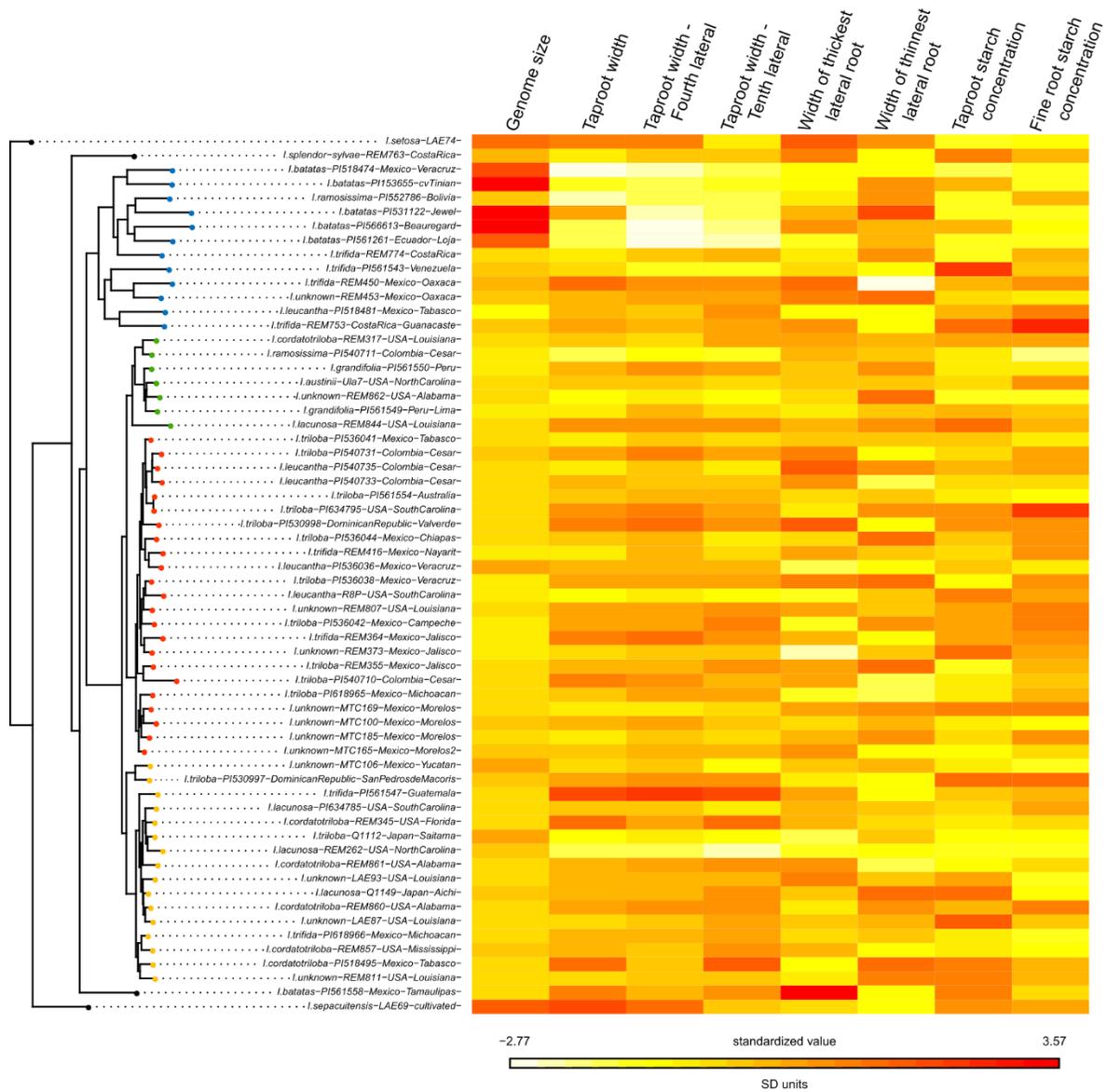


Figure 4.7. Phylogeny of the Batatas complex with genome size and root trait values. Trait values are displayed in the heatmap. Each column is standardized to be able to compare across columns. Shown to the left is the concatenation tree. Colored dots on tips of the phylogeny correspond to clade colors in Figure 4.2.

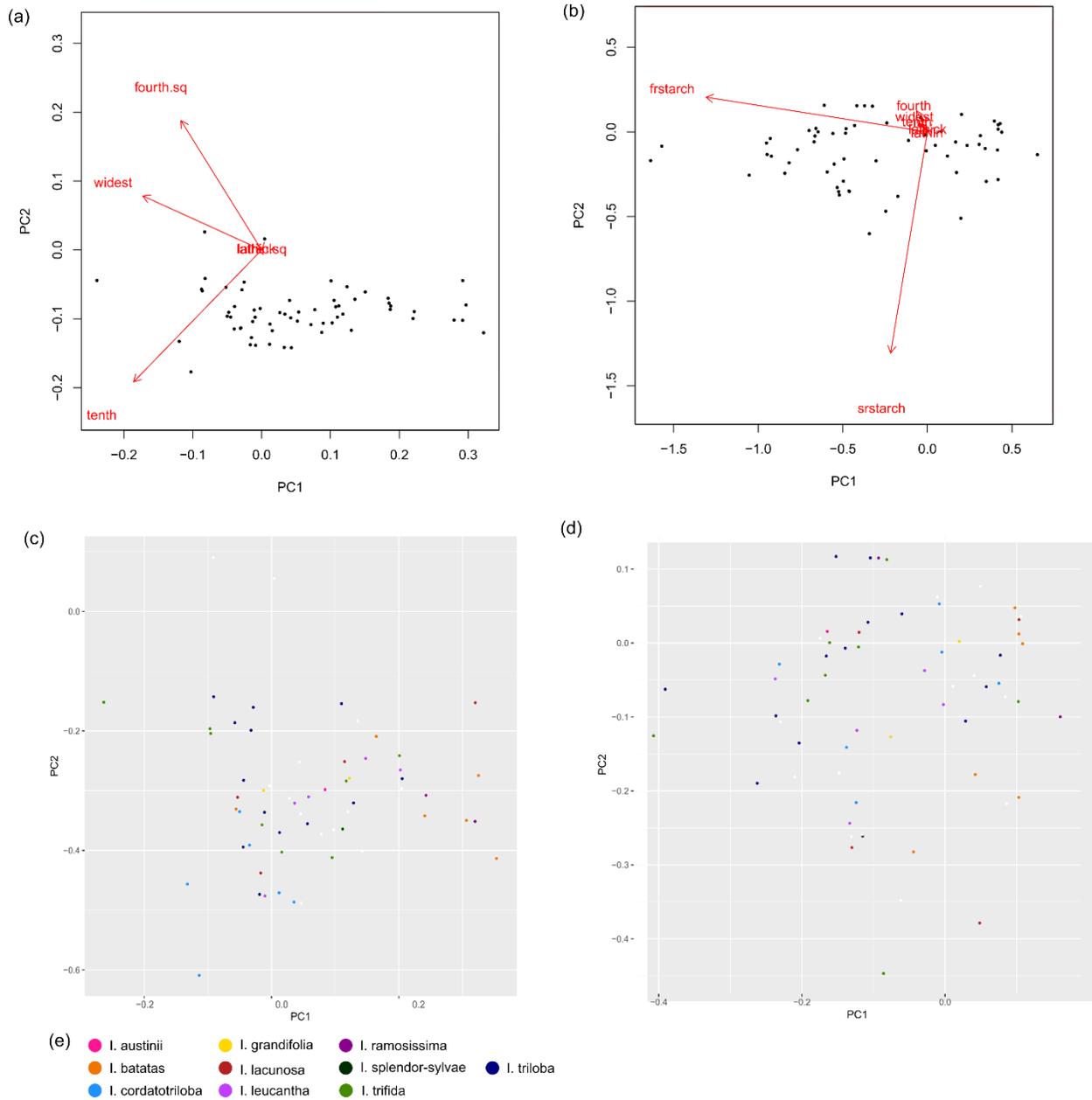


Figure 4.8. Biplot of phylogenetic principal components analyses for the dataset which did not include starch (a, c) and the dataset that did include starch measurements (b, d). Colors in (c) and (d) correspond to species designations and are denoted in (e). White dots in (c) and (d) denote individuals which could not be identified to species.

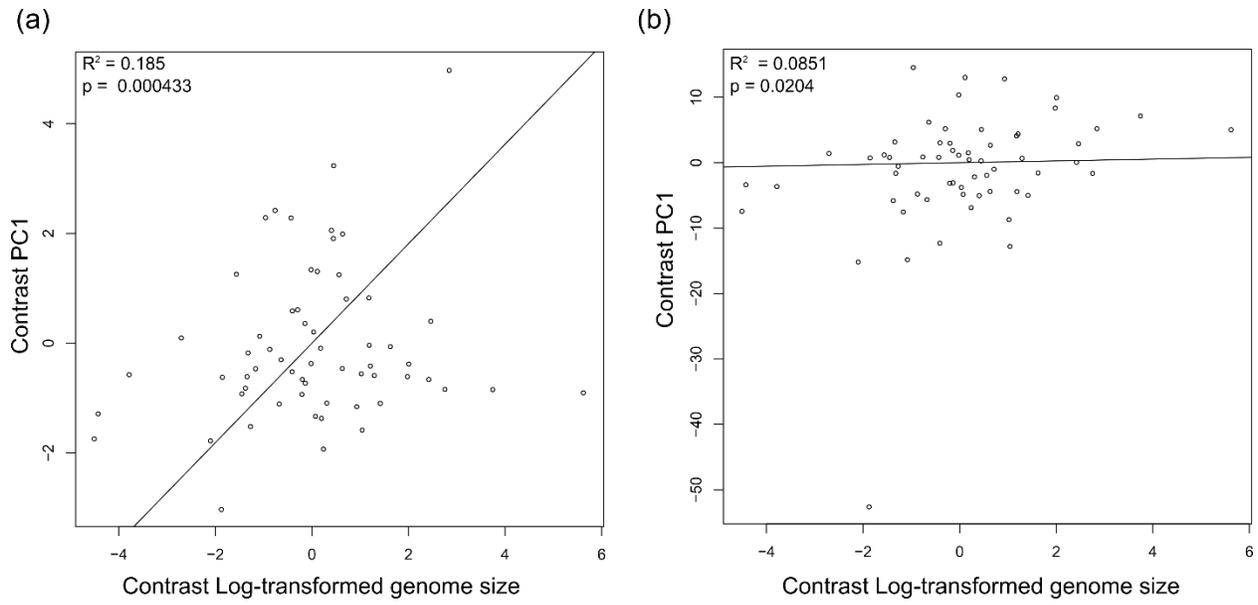


Figure 4.9. Linear regression between phylogenetic independent contrasts of the first PC axis (PC1) and log-transformed genome size in datasets which did not include starch (a) and did include starch (b) in the phylogenetic principal components analysis.

CHAPTER V:

PARALLEL EVOLUTION OF STORAGE ROOTS IN MORNING GLORIES (CONVOLVULACEAE)<sup>4</sup>

<sup>4</sup>Eserman, Lauren A., Robert L. Jarret, Jim H. Leebens-Mack. Submitted to *New Phytologist*, July 12, 2017.

## **Abstract**

Storage roots are an ecologically and agriculturally important plant trait that have evolved numerous times in angiosperms. Storage roots primarily function to store carbohydrates underground as reserves for perennial species. In morning glories, storage roots are well characterized in the crop species sweetpotato, where starch is stored in anomalous cambium. This anomalous cambium proliferates, and roots thicken to accommodate the addition of starch storage tissue. In morning glories, storage roots have evolved numerous times. The primary goal of this study is to understand whether this was through parallel evolution, where species use a common genetic mechanism to achieve storage root formation, or through convergent evolution, where storage roots in distantly related species are formed using a different set of genes. Pairs of species where one forms storage roots and the other does not were sampled from two tribes in the morning glory family, the Ipomoeae and Merremieae. Root anatomy in storage roots and fine roots was examined. Furthermore, we sequenced total mRNA from storage roots and fine roots in these species and analyzed differential gene expression. Anatomical results reveal that storage roots of species in the Ipomoeae tribe, such as sweetpotato, accumulate starch differently than species in the Merremieae tribe. In both storage root forming species, more genes were found to be upregulated in storage roots compared to fine roots. Further, we find that fifty-seven orthologous genes were differentially expressed between storage roots and fine roots in both storage root forming species. These genes are primarily involved in starch biosynthesis, regulation of starch biosynthesis, and transcription factor activity. Taken together, these results demonstrate that storage roots of species from both morning glory tribes are anatomically different but utilize a common core set of genes in storage root formation. This is consistent with

a pattern of parallel evolution. In the future, a time course experiment would be beneficial to characterize the fine scale changes associated with storage root formation.

## **Introduction**

Parallel and convergent evolution of complex morphological traits has long been of interest to evolutionary biologists, who have noted that functionally and morphologically similar phenotypes have evolved independently in unrelated lineages . Studies characterizing the genetic basis of independent phenotypic evolution have concluded that many traits evolve convergently, appearing functionally similar but utilizing different genetic mechanisms. Alternatively, traits evolving in parallel have the same genetic basis (Haas and Simpson 1946; Scotland 2011). Often, differentiating between these alternative evolutionary scenarios is difficult. Studies comparing morphology, anatomy, gene expression and other aspects of a trait can provide insights into whether a trait evolved convergently or in parallel.

Morning glories offer an ideal system in which to address hypotheses regarding convergent versus parallel evolution. In morning glories, storage root formation has been either lost or gained at least ten times independently; however, it is unclear whether the ancestor of all morning glories was able to form storage roots (Eserman et al. 2014). Studies characterizing storage root development in sweetpotato have demonstrated that a storage root is simply a modification of the taproot, an adventitious root, and/or one or more lateral roots such that the root cambium expands and the starch-storage tissue, termed anomalous cambium, proliferates (Artschwager 1924; Wilson and Lowe 1973; Lowe and Wilson 1974a, 1974b; Firon et al. 2013). The proliferation of anomalous cambium in the root cambium expands the root so that storage roots are much greater in diameter than roots which do not function in long-term starch storage. Studies analyzing gene expression differences between fine and storage roots in *Ipomoea batatas*

(sweetpotato) found that genes in the starch biosynthesis are highly expressed and lignin biosynthesis genes are lowly expressed in storage roots compared to fine roots (Firon et al. 2013). Studies have also implicated three genes in the development of storage roots, two of which are MADS-box transcription factors (Ku et al. 2008; Noh et al. 2010) and the other is an alpha-expansin gene (Noh et al. 2013). However, these studies were strictly limited to sweetpotato. Comparative studies may reveal genes involved in storage root formation across distantly related species

In addition to the evolutionary importance, storage roots have economic and ecological significance as well. Sweetpotato [*Ipomoea batatas* (L.) Lam.] ranks among the ten most important crop species for human nutrition. In 2014, over 100 million tonnes of sweetpotato were produced worldwide (FAO 2016). The large storage roots are an important source of carbohydrates and vitamin A in developing countries (Hotz et al. 2012). More generally, storage roots play a key role in the life history and ecological strategies of plants, as perennial species tend to mobilize starch to roots year-round and thus form storage roots but annual species cease starch mobilization after only a few months (De Souza and Viera Da Silva 1987). Additionally, root carbohydrate reserves are necessary for resprouting after cutting or large-scale events such as fire (Bowen and Pate 1993; Bell et al. 1996; Vriet et al. 2014).

Given what is known about the developmental biology and anatomy of storage roots, lineages that form storage roots may represent instances of either convergent or parallel evolution. In this study, we aim to: 1) to understand the anatomy of storage roots in morning glories and 2) to characterize gene expression during an early stage of storage root formation. If we observe that storage roots from distantly related morning glory lineages are anatomically similar and share an overlapping set of differentially expressed orthologous genes, this would

provide evidence supporting the hypothesis that storage roots evolved prior to the diversification of morning glories and were subsequently lost in lineages that do not form storage roots (parallel evolution). However, if we observe that storage roots are anatomically dissimilar and share few to no differentially expressed orthologous genes, this would support the hypothesis that storage roots evolved independently in storage root forming lineages (convergent evolution). Using this comparative approach, we can better understand the genetic mechanisms and evolutionary origins of storage root formation. Through this work we ultimately seek to understand the genetic basis of storage root formation and whether independent lineages utilize the same or different genetic mechanisms during storage root development across the morning glory phylogeny.

## **Methods**

### *Plant material*

Three pairs of closely-related species were selected from across the morning glory phylogeny, where one member of the species pair produces storage roots and fine roots and the other produces only fine roots. The three storage root forming species are *Ipomoea batatas* (L.) Lam. (sweetpotato), *I. lindheimeri* A. Gray and *Merremia dissecta* (Jacq.) Hallier f., and the species that produce only fine roots are *I. trifida* G. Don, *I. nil* (L.) Roth, and *M. quinquefolia* (L.) Hallier f. All three pairs of species were utilized for anatomical observations. Four species, *I. batatas*, *I. trifida*, *M. dissecta* and *M. quinquefolia*, were used for transcriptome sequencing so that we could directly contrast gene expression of different observed root architectures. Three biological replicates were chosen for each species except *I. trifida*, where RNA-seq libraries for one sample consistently failed.

Sweetpotato is vegetatively propagated, so cuttings were planted of the three cultivars with three true leaves. Seeds of the other five species were scarified before planting. Seeds and cuttings were planted in Fafard 3B mix in 4" square pots. Seeds were allowed to germinate for 1 week in the UGA Greenhouses. Plants were then moved to a growth chamber under an 8 hour photoperiod and 30°/25°C day/ night temperatures (Ku et al. 2008). Previous studies have found that storage root formation occurs within four to six weeks after planting in sweetpotato (Nakatani et al. 2002; Firon et al. 2013); therefore, plants in this study were grown for six weeks prior to sampling. Roots were sampled using the following procedure: roots were removed from dirt, washed in tap water, and rinsed a final time in nuclease-free molecular biology grade water. The primary root was dissected from the whole plant, and fine lateral roots were then dissected from the primary root. Fresh root tissue was flash frozen in liquid nitrogen and were subsequently stored at -80°C until RNA isolation. Alternatively, fresh root tissue was used immediately for anatomical observations.

#### *Anatomical observations*

Fresh root tissue was sectioned by hand with a razor blade. A main goal of this was to observe the spatial deposition of starch in cross sections of the root; therefore, fresh sections were necessary because starch is removed during standard tissue clearing (Jensen 1962). Serial sections were taken from fine roots and from two places on the taproot or storage root: 1) after the 4<sup>th</sup> lateral root, and 2) after the 10<sup>th</sup> lateral root. Sections were stained with Lugol's iodine, a solution of iodine and potassium iodide which indicates the presence of starch, or phloroglucinol-HCl, which stains lignin (Turrell and Fisher 1942), immediately following sectioning. Stained sections were mounted in a filtered 20% CaCl<sub>2</sub> solution (Herr 1992). Mounted sections were viewed with a Zeiss Axio microscope with attached camera under either

a 2.5x or 10x objective lens. Sections too large to be viewed in a single field of vision using the 2.5x objective lens were captured in multiple images which were then stitched together using the image stitching plugin for the Fiji distribution of ImageJ (Preibisch et al. 2009; Schindelin et al. 2012, 2015). Field of vision length was determined using a standard microscope scale, and scale bars were added to image in ImageJ.

#### *RNA isolations and library construction*

Total RNA was isolated from frozen root tissue using the standard Trizol protocol (Life Technologies). RNA was eluted in molecular biology grade H<sub>2</sub>O following isolation. DNA was removed using the TURBO DNA-free kit (Thermo Fisher Scientific). Prior to library construction, RNA quality was assessed with the Agilent Bioanalyzer 2100 using the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA). mRNA was isolated from total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Inc.). The first mRNA isolations performed using the recommended total RNA input yielded low mRNA concentrations. Therefore, the amount of total RNA added to the mRNA isolation protocol was increased to 5 µg, the maximum recommended RNA input. Libraries were constructed with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Inc.) using the standard protocol with slight modifications. Libraries were amplified with 15 PCR cycles. An initial test set of libraries showed adapter dimer peaks; therefore, the adapter was diluted 1.25 µM rather than the standard 1.5 µM, which eliminated adapter dimer peaks in future libraries. The library preparation protocol used in this experiment implements the dUTP method (Parkhomchuk et al. 2009) to generate stranded libraries.

Libraries were quantified using quantitative real-time PCR prior to sequencing. Libraries were diluted to 10 nM for sequencing. Barcoded and diluted libraries were pooled before

sequencing. All libraries were sequenced at the Georgia Genomics Facility on the Illumina NextSeq platform with paired-end 150 bp reads.

### *Transcriptome analysis*

Reads for each species were assembled separately into transcripts with the Trinity software suite version r20140717 (Grabherr et al. 2011). Within-species transcriptome assembly and analysis followed the developed Trinity pipeline (Haas et al., 2013). Read quality was assessed with FastQC. Prior to assembly, reads were quality trimmed with Trimmomatic as implemented in the Trinity package. Bases at the beginning and end of a read with a phred score less than 5 were removed. In addition, reads less than 50 bp long were removed. Reads for each library were digitally normalized to a maximum of 50x coverage within Trinity (--normalize\_reads) to accelerate the assembly process. Reads were considered paired-end in the assembly, where the first read of the pair was considered the reverse read and the second read was the forward read (--SS\_lib\_type RF).

We then filtered assemblies to remove lowly supported isoforms and contaminants. We used RSEM version 1.2.20 (Li and Dewey 2011) to estimate gene and transcript abundances as implemented in the Trinity package (align\_and\_estimate\_abundance.pl script). Non-normalized reads were mapped to each transcriptome assembly with Bowtie 2 (Langmead and Salzberg 2012). Isoforms which were supported by less than 30% of the total reads for a gene from two or more biological replicates or had an FPKM less than 2 were removed, as these represent possible assembly artifacts. Filtering was performed using the perl script filter\_fasta\_by\_rsem\_values.pl in the Trinity software package (Haas et al. 2013). To remove contaminants, we annotated the assembled transcriptomes in Trinotate (Haas et al. 2013) using a blastx of the filtered assembly against the Uniprot database. Transcripts with annotations from any taxon other than

Viridiplantae with an e-value greater  $1e^{-5}$  and 40% identity were removed as potential contaminants. Finally, the program DeconSeq version 0.4.2 (Schmieder and Edwards 2011) was used further filter any remaining bacterial, viral, and human contaminant sequences.

RSEM (Li and Dewey 2011) and Bowtie 2 (Langmead and Salzberg 2012) were again used to map reads from individual libraries back to the filtered transcriptome assemblies and calculate transcript abundances. EdgeR (Robinson et al. 2010) was then used to assess differentially expressed genes between storage roots and fine roots of *Ipomoea batatas* and *Merremia dissecta* using perl scripts from the Trinity analysis pipeline (Haas et al. 2013). EdgeR was run separately for each species and incorporated biological replicates for each tissue type. FPKM values for each library were normalized by library size. This normalization process is referred to as “Trimmed Mean of M-values”, or TMM, normalization (Robinson and Oshlack 2010). Only TMM-normalized FPKM values were used for differential expression analysis. Transcripts were considered significantly differentially expressed at a false discovery rate (FDR) less than 0.05 and a log fold change of 2. We then generated Euclidean distances among transcripts and libraries and used a complete linkage clustering approach on the Euclidean distance matrices to cluster transcripts and libraries in edgeR.

Protein coding regions were identified from the final filtered assemblies using the program Transdecoder (Haas et al. 2013). Protein sequences shorter than 50 amino acid residues long were not kept in the final set of peptide sequences. Functional annotation utilized the standard Trinotate pipeline (Haas et al. 2013), which incorporated a blastx search of the assembled transcripts against the Uniprot database and a blastp search of the peptide sequences inferred from Transdecoder against the Uniprot database. These results as well as gene ontology

(GO) term annotations of the best gene match in Uniprot were incorporated into a SQLite database using Trinotate (Haas et al. 2013).

Peptide sequences from the final filtered assemblies from all four species were sorted into gene families with OrthoFinder (Emms and Kelly 2015) to determine orthology among transcripts from the four species. Coding sequences of the gene families estimated from OrthoFinder were aligned in SATé-II (Liu et al. 2012). Gene trees were estimated in RAxML (Stamatakis 2014), and node support was determined using 500 bootstrap replicates.

## Results

### *Root anatomy*

Results of the root anatomical observations are shown in Figure 5.1. There were three main results from this. First, fine roots of all six species are anatomically similar and exhibit the typical eudicot root anatomy with a highly organized vascular cambium in the center and a larger cortex. Second, we found that the taproot of the species that do not form storage roots appear similar, accumulate very little starch, and do not show evidence of proliferation of anomalous cambium. Third, storage roots of the three storage root forming species showed similar starch accumulation, specifically, proliferation of the starch-accumulating anomalous cambium that occurred within the bounds of the endodermis. Finally, the vascular tissue in storage roots of *Ipomoea batatas* and *I. lindheimeri* appeared visually similar, where the starch-accumulating anomalous cambium disrupted the organization of the vascular tissue. In contrast, the vascular tissue of storage roots of *Merremia dissecta* appeared markedly different such that vascular tissue was tightly organized in the center of the cross section.

### *Transcriptome assembly statistics*

The final dataset included eighteen RNAseq libraries from two pairs of morning glory species. Transcriptome assembly statistics are shown in Table 5.1. Before filtering, the *Merremia quinquefolia* transcriptome had the largest number of transcripts, and the *I. trifida* transcriptome had the fewest assembled transcripts. Transcript N50 ranged from 952 to 1277 nt. We then filtered the raw assemblies by isoform percentage and FPKM, which resulted in a 42-70% reduction in the number of transcripts in the assembly (Table 5.2). This step removed potentially erroneous transcripts that were not supported by re-mapped reads. Further filtering of bacterial, fungal, algal, and viral transcripts using Swiss-prot annotations and DeconSeq resulted in an additional ca. 3900-5700 transcripts removed from each assembly. Only the transcriptomes filtered by isoform percentage and FPKM and which had contaminants removed were used for downstream analyses.

### *Within species differential gene expression*

We assessed differential gene expression between storage roots and fine roots in *Ipomoea batatas* and *Merremia dissecta* separately. After accounting for multiple comparisons, there were 2643 genes DE between storage roots and fine roots in *I. batatas* and 219 DE genes in *M. dissecta* at a FDR <0.05 (Figure 5.2a,b). In both species, there were more transcripts highly expressed in storage roots than in fine roots. As a convention, upregulated transcripts refers to those more highly expressed in storage roots vs. fine roots and downregulated refers to transcripts lowly expressed in storage roots compared to fine roots. In *I. batatas*, 1642 transcripts were upregulated and 1001 transcripts were downregulated. In *Merremia dissecta*, there were 178 upregulated transcripts and 41 downregulated transcripts.

The top ten most abundant gene ontology annotations for the differentially expressed genes in *I. batatas* and *M. dissecta* are found in Table 5.3. When we compare the ten most abundant GO annotations from genes DE in *I. batatas* and *M. dissecta*, we find that eight of these GO terms overlap. Additionally, many of the most enriched GO terms were involved in transcription or are annotated as having transcription factor activity (Table 5.3).

#### *Between species differential gene expression*

To compare gene expression between orthologs of different species, we sorted transcripts into orthologous groups with OrthoFinder (Emms and Kelly 2015). We then queried the orthologous groups for known set of transcripts differentially expressed (DE) between storage and fine roots in *Ipomoea batatas* and *Merremia dissecta*. We found there were 57 orthologous genes DE between storage roots and fine roots of both species (Figure 5.2c). We then examined GO term annotations for the set of orthologous DE transcripts (Table 5.4). Transcripts annotated with amyloplast or starch biosynthetic activity were found to represent a larger percent of the total GO annotations in the set of shared DE transcripts than in the DE transcripts from *I. batatas* and *M. dissecta* analyzed separately (Tables 5.3, 5.4). Similarly, we examined the functional annotation of these transcripts and found that some of these DE genes share close homology with transcription factors, alpha-expansin genes, genes that function in the starch biosynthetic pathway, and one that functions in the starch degradation pathway.

#### *Among species differential gene expression*

We then wanted to examine the expression of genes in the starch biosynthetic pathway (Figure 5.3). Most genes in the starch biosynthesis pathway were found to be lowly expressed. However, orthologs of GLGL1 and SSG1 were significantly differentially expressed in *Ipomoea batatas* and *Merremia dissecta* (Figure 5.3). These genes were highly expressed in storage roots

and lowly expressed in fine roots, except for GLGL1 in *M. quinquefolia* (Figure 5.3). Furthermore, we examined expression of transcripts annotated as having transcription factor activity, where orthologs were differentially expressed in both *I. batatas* and *M. dissecta* (Figure 5.4). In all cases, orthologs of the shared differentially expressed transcription factors were more highly expressed in storage roots than fine roots (Figure 5.4).

## Discussion

### *Root anatomy*

Results of the root anatomical work clearly show that the storage roots of species in the tribe Ipomoeae (*Ipomoea batatas* and *I. lindheimeri*) are anatomically quite different from storage roots of *Merremia dissecta*, a member of the sister tribe Merremieae. Anomalous cambium proliferation occurred in all three storage root forming species; however, xylem organization differed greatly in *M. dissecta* compared to storage roots of the other two species (Figure 5.1). Our findings are consistent with other studies examining root anatomical structure of sweetpotato (Artschwager 1924; Wilson and Lowe 1973; Lowe and Wilson 1974a, 1974b; Firon et al. 2013). However, we had no a priori expectations with regard to root anatomy of all other species included in this study, as this is the first to document root anatomy of *I. lindheimeri*, *I. nil*, *I. trifida*, *M. dissecta*, and *M. quinquefolia*.

### *Comparison of gene expression in all species*

Based on the anatomical results, we can generate expectations with respect to the transcriptome experiment. Starch accumulation and anomalous cambium proliferation occurred similarly in storage roots of all three species; however, xylem organization was quite different in storage roots of *M. dissecta*. Therefore, it is likely that genes involved in starch biosynthesis and

cell proliferation will be differentially expressed between storage and fine roots in both species, but genes involved in xylem organization may not show the same gene expression patterns between species.

At a broad level, we found more genes were found to be upregulated in storage roots compared to fine roots in both *I. batatas* and *M. dissecta*. Interestingly, this result is in contrast to a previous RNAseq study in sweetpotato which found an approximately equal number of genes up- and downregulated in storage roots compared to fine roots (Firon et al. 2013). We sampled roots six weeks after planting in contrast to Firon et al. (2013), which sampled roots at four weeks. Given that we are sampling at a slightly later growth stage, perhaps we are capturing a more active stage of storage root bulking in this study. In the future, closer examination of the anatomical and gene expression changes during the very early stages of storage root formation would provide further insights into the development of this trait.

#### *Starch biosynthetic pathway*

Starch biosynthesis occurs as part of a complex and dynamic pathway and the enzymes and transport proteins involved depend heavily upon the tissue in which starch is being synthesized. The process differs in photosynthetic and heterotrophic tissues (Bahaji et al. 2014). Therefore, we focused on the starch pathway that has been characterized in potato tubers from Bahaji et al. (2014) because it is the most well-characterized starch biosynthetic pathway in heterotrophic tissue in a species closely related to sweetpotato.

Whereas in photosynthetic tissue sucrose is broken into fructose and glucose prior to starch synthesis, in heterotrophic tissues, sucrose is directly converted to UDP-glucose before starch biosynthesis (Bahaji et al. 2014). In addition, the downstream conversion of UDP-glucose to starch intermediates differs between eudicot and monocot heterotrophic tissues. UDP-glucose

is converted to glucose-1-phosphate by the enzyme UDP-glucose pyrophosphorylase (UGPA). Glucose-1-phosphate is then either transported from the cytosol into the amyloplast or converted in the cytosol to glucose-6-phosphate by the enzyme phosphoglucomutase (PGMP). Glucose-6-phosphate is then transported into the amyloplast by the transport protein glucose-6-phosphate translocator (GPT) where it is converted back to glucose-1-phosphate by PGMP. Glucose-1-phosphate is converted to ADP-glucose by the action of ADP-glucose pyrophosphorylase (GLGL), which requires an input of ATP. ADP-glucose is then converted to the main components of starch by granule-bound starch synthase (SSG) to generate amylose or starch synthase (SSY) and starch branching enzymes (GLGB) to generate amylopectin.

In the context of this study, we found that orthologs of two genes involved in starch biosynthesis had significantly higher expression in storage roots compared to fine roots in both *Ipomoea batatas* and *Merremia dissecta* (Figure 5.3). In this study, GLGL1 and SSG1 were significantly differentially expressed between storage roots and fine roots of sweetpotato and *M. dissecta* (Figure 5.3). GLGL acts downstream in the pathway, directly upstream of SSG, which is involved in the synthesis of amylose (Bahaji et al. 2014). Generally, amylose content in sweetpotato cultivars is high, ranging from 20-33% of total starch content (Walter et al. 2000; Waramboi et al. 2011), much higher than in other starch-rich root and tuber crops such as cassava (Mejia-Aguero et al. 2012).

This examination must be taken with the caveat that starch accumulation and bulking may occur through different mechanisms in sweetpotato and potato. First, sweetpotato storage roots and potato tubers arise from different tissue types; storage roots from root tissue and tubers from stem tissue (Xu et al. 2011). Furthermore, tuber formation in potato is controlled by a homologue of flowering locus T (SP6A), and the process of tuber initiation is dependent on

photoperiod (Xu et al. 2011). However, experimental evidence has demonstrated that sweetpotato storage root initiation occurs under both long and short day regimes (Loretan et al. 1994). Future functional genomic research involving sweetpotato and its close relatives is necessary to elucidate the exact mechanisms of starch biosynthesis and storage.

### *Transcription factors*

Of the fifty-seven orthologous genes differentially expressed between storage roots and fine roots, seven were annotated as having transcription factor activity (Figure 5.4). When we further examine the annotated functions of these genes, two stand out as potential candidate regulators of storage root formation.

IDD5, also called RAVEN, has been shown to positively regulate starch synthase in *Arabidopsis thaliana* (Ingkasuwan et al. 2012). Additionally, IDD5 is part of a larger regulatory network that, among other functions, regulates spatial patterning of root tissue through asymmetric cell division (Welch et al. 2007; Hassan et al. 2010; Ingkasuwan et al. 2012). Many members of the larger regulatory network to which IDD5 belongs were found to be differentially expressed between SRs and FRs in sweetpotato cv. Georgia Jet and Xushu (Tao et al. 2012; Firon et al. 2013), suggesting a possible role of IDD5 and members of this regulatory network in storage root formation.

Similarly, WOX4 orthologs were DE between storage roots and fine roots of both *I. batatas* and *M. dissecta*. This gene has been shown to play a critical role in vasculature proliferation and secondary growth in *Arabidopsis thaliana*, and functions specifically within the cambium of stems and roots (Suer et al. 2011; EtcHELLS et al. 2013). Perhaps this gene plays a role in the anomalous cambium proliferation that we observe in storage roots of *I. batatas* and *M. dissecta*.

## Conclusion

The anatomical results suggested that storage roots differ from fine roots in starch content, deposition and vasculature patterning. As expected, we found significantly higher expression of genes involved directly in starch biosynthesis in both storage root forming species and increased expression of *IDD5*, a transcription factor known to regulate starch biosynthesis in *Arabidopsis* (Ingkasuwan et al. 2012). Similarly, we found significant upregulation of *WOX4*, a gene known to be involved in vasculature proliferation in *Arabidopsis* (Suer et al. 2011; Etchells et al. 2013). Given the large number of orthologous genes DE between storage roots and fine roots, we hypothesize that there was a single origin of storage roots before the divergence of the morning glory tribes Ipomoeae and Merremieae given that storage roots in the species examined are superficially anatomically different but store starch similarly through anomalous cambium proliferation. To further support this hypothesis, we find that many of the same genes were differentially expressed between storage roots and fine roots in sweetpotato and *Merremia dissecta*. However, an alternative hypothesis, that storage roots evolved multiple times independently using the same genetic mechanisms, cannot be directly rejected by these results. Therefore, much more work must be done to test these hypotheses in a rigorous framework. The findings presented here present a first step in understanding the evolution and development of a plant trait that has received little attention to date but is economically and ecologically important. These results further demonstrate the power of comparative studies to understand the development of a trait and its evolution in a deeper way than to examine a single species.

## References

- Artschwager, E. 1924. On the anatomy of the sweet potato with notes on internal breakdown. J. Agric. Res. 27:157–166.

- Bahaji, A., J. Li, Á. M. Sánchez-López, E. Baroja-Fernández, F. J. Muñoz, M. Ovecka, G. Almagro, M. Montero, I. Ezquer, E. Etxeberria, and J. Pozueta-Romero. 2014. Starch biosynthesis, its regulation and biotechnological approaches to improve crop yields. *Biotechnol. Adv.* 32:87–106.
- Bell, T. L., J. S. Pate, and K. W. Dixon. 1996. Relationships Between Fire Response, Morphology, Root Anatomy and Starch Distribution in South-west Australian Epacridaceae. *Ann. Bot.* 77:357–364.
- Bowen, B. J., and J. S. Pate. 1993. The significance of root starch in post-fire shoot recovery of the resprouter *Stirlingia latifolia* R. Br. (Proteaceae). *Ann. Bot.* 72:7–16.
- De Souza, J. G., and J. Viera Da Silva. 1987. Partitioning of carbohydrates in annual and perennial cotton (*Gossypium hirsutum* L.). *J. Exp. Bot.* 38:1211–1218.
- Emms, D. M., and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.
- Eserman, L. A., G. P. Tiley, R. L. Jarret, J. H. Leebens-Mack, and R. E. Miller. 2014. Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* 101:92–103.
- Etchells, J. P., C. M. Provost, L. Mishra, and S. R. Turner. 2013. WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. *Development* 140:2224–2234.
- FAO. 2016. Food and Agriculture Organization of the United Nations.
- Firon, N., D. LaBonte, A. Villordon, Y. Kfir, J. Solis, E. Lapis, T. S. Perlman, A. Doron-Faigenboim, A. Hetzroni, L. Althan, and L. Adani Nadir. 2013. Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics* 14:460.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, D. Philip, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.* 8:1–43.

- Haas, O., and G. G. Simpson. 1946. Analysis of Some Phylogenetic Terms , with Attempts at Redefinition. *Proc. Am. Philos. Soc.* 90:319–349.
- Hassan, H., B. Scheres, and I. Blilou. 2010. JACKDAW controls epidermal patterning in the *Arabidopsis* root meristem through a non-cell-autonomous mechanism. *Development* 137:1523–1529.
- Herr, J. M. 1992. New uses for calcium chloride solution as a mounting medium. *Biotech. Histochem.* 67:9–13.
- Hotz, C., C. Loechl, A. de Brauw, P. Eozenou, D. Gilligan, M. Moursi, B. Munhaua, P. van Jaarsveld, A. Carriquiry, and J. V. Meenakshi. 2012. A large-scale intervention to introduce orange sweet potato in rural Mozambique increases vitamin A intakes among children and women. *Br. J. Nutr.* 108:163–76.
- Ingkasuwan, P., S. Netrphan, S. Prasitwattanaseree, M. Tanticharoen, S. Bhumiratana, A. Meechai, J. Chaijaruwanich, H. Takahashi, and S. Cheevadhanarak. 2012. Inferring transcriptional gene regulation network of starch metabolism in *Arabidopsis thaliana* leaves using graphical gaussian model. *BMC Syst. Biol.* 6:100.
- Jensen, W. A. 1962. *Botanical histochemistry: principles and practice.* W. H. Freeman, San Francisco.
- Ku, A. T., Y.-S. Huang, Y.-S. Wang, D. Ma, and K.-W. Yeh. 2008. IbMADS1 (*Ipomoea batatas* MADS-box 1 gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). *Ann. Bot.* 102:57–67.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–9.
- Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Liu, K., T. J. Warnow, M. T. Holder, S. M. Nelesen, J. Yu, A. P. Stamatakis, and C. R. Linder. 2012. SATE-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Syst. Biol.* 61:90–106.
- Loretan, P. A., C. K. Bonsi, D. G. Mortley, R. M. Wheeler, C. L. Mackowiak, W. A. Hill, C. E. Morris, A. A. Trotman, and P. P. David. 1994. Effects of several environmental factors on sweetpotato growth. *Adv. Sp. Res.* 14:277–80.
- Lowe, S. B., and L. A. Wilson. 1974a. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 1. Tuber initiation, tuber growth and partition of assimilate. *Ann. Bot.* 38:307–317.
- Lowe, S. B., and L. A. Wilson. 1974b. Comparative analysis of tuber development in six sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. 2. Interrelationships between tuber shape and yield. *Ann. Bot.* 38:319–326.

- Mejia-Aguero, L. E., F. Galeno, O. Hernandez-Hernandez, J. Matehus, and J. Tovar. 2012. Starch determination, amylose content and susceptibility to in vitro amylolysis in flours from the roots of 25 cassava varieties. *J. Sci. Food Agric.* 92:673–678.
- Nakatani, M., M. Tanaka, and M. Yoshinaga. 2002. Physiological and Anatomical Characterization of a Late-storage Root-forming Mutant of Sweetpotato. *J. Am. Soc. Hortic. Sci.* 127:178–183.
- Noh, S. A., H.-S. Lee, E. J. Huh, G. H. Huh, K.-H. Paek, J. S. Shin, and J. M. Bae. 2010. SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (*Ipomoea batatas*). *J. Exp. Bot.* 61:1337–49.
- Noh, S. A., H.-S. Lee, Y.-S. Kim, K.-H. Paek, J. S. Shin, and J. M. Bae. 2013. Down-regulation of the IbEXP1 gene enhanced storage root development in sweetpotato. *J. Exp. Bot.* 64:129–142.
- Parkhomchuk, D., T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, and A. Soldatov. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37:e123.
- Preibisch, S., S. Saalfeld, and P. Tomancak. 2009. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* 25:1463–5.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Robinson, M., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25.
- Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9:676–682.
- Schindelin, J., C. T. Rueden, M. C. Hiner, and K. W. Eliceiri. 2015. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* 82:518–529.
- Schmieder, R., and R. Edwards. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6.
- Scotland, R. W. 2011. What is parallelism? *Evol. Dev.* 13:214–27.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

- Suer, S., J. Agusti, P. Sanchez, M. Schwarz, and T. Greb. 2011. WOX4 imparts auxin responsiveness to cambium cells in *Arabidopsis*. *Plant Cell* 23:3247–59.
- Tao, X., Y.-H. Gu, H.-Y. Wang, W. Zheng, X. Li, C.-W. Zhao, and Y.-Z. Zhang. 2012. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam]. *PLoS One* 7:e36234.
- Turrell, F. M., and P. L. Fisher. 1942. The proximate chemical constituents of *Citrus* woods, with special reference to lignin. *Plant Physiol.* 17:558–581.
- Vriet, C., A. M. Smith, and T. L. Wang. 2014. Root starch reserves are necessary for vigorous re-growth following cutting back in *Lotus japonicus*. *PLoS One* 9:1–7.
- Walter, W. M., V. D. Truong, D. P. Wiesenborn, and P. Carvajal. 2000. Rheological and physicochemical properties of starches from moist- and dry-type sweetpotatoes. *J. Agric. Food Chem.* 48:2937–2942.
- Waramboi, J. G., S. Dennien, M. J. Gidley, and P. A. Sopade. 2011. Characterisation of sweetpotato from Papua New Guinea and Australia: Physicochemical, pasting and gelatinisation properties. *Food Chem.* 126:1759–1770.
- Welch, D., H. Hassan, I. Blilou, R. Immink, R. Heidstra, and B. Scheres. 2007. *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.* 21:2196–2204.
- Wilson, L. A., and S. B. Lowe. 1973. The anatomy of the root system in West Indian sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars. *Ann. Bot.* 37:633–643.
- Xu, X., S. Pan, S. Cheng, B. Zhang, D. Mu, P. Ni, G. Zhang, S. Yang, R. Li, J. Wang, G. Orjeda, F. Guzman, M. Torres, R. Lozano, O. Ponce, D. Martinez, G. De la Cruz, S. K. Chakrabarti, V. U. Patil, K. G. Skryabin, B. B. Kuznetsov, N. V. Ravin, T. V. Kolganova, A. V. Beletsky, A. V. Mardanov, A. Di Genova, D. M. Bolser, D. M. a Martin, G. Li, Y. Yang, H. Kuang, Q. Hu, X. Xiong, G. J. Bishop, B. Sagredo, N. Mejía, W. Zagorski, R. Gromadka, J. Gawor, P. Szczesny, S. Huang, Z. Zhang, C. Liang, J. He, Y. Li, Y. He, J. Xu, Y. Zhang, B. Xie, Y. Du, D. Qu, M. Bonierbale, M. Ghislain, M. D. R. Herrera, G. Giuliano, M. Pietrella, G. Perrotta, P. Facella, K. O'Brien, S. E. Feingold, L. E. Barreiro, G. a Massa, L. Diambra, B. R. Whitty, B. Vaillancourt, H. Lin, A. N. Massa, M. Geoffroy, S. Lundback, D. DellaPenna, C. R. Buell, S. K. Sharma, D. F. Marshall, R. Waugh, G. J. Bryan, M. Destefanis, I. Nagy, D. Milbourne, S. J. Thomson, M. Fiers, J. M. E. Jacobs, K. L. Nielsen, M. Sønderkær, M. Iovene, G. a Torres, J. Jiang, R. E. Veilleux, C. W. B. Bachem, J. de Boer, T. Borm, B. Kloosterman, H. van Eck, E. Datema, B. T. L. Hekkert, A. Goverse, R. C. H. J. van Ham, and R. G. F. Visser. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195.

Table 5.1. Transcriptome assembly statistics.

	<i>I. batatas</i>	<i>I. trifida</i>	<i>M. dissecta</i>	<i>M. quinquefolia</i>
<b>Total reads (PE 150bp)</b>	39,632,572	15,657,942	44,877,650	64,267,290
<b>No. of transcripts</b>	245140	119153	254174	363820
<b>%GC</b>	40.97	42.1	39.67	39.37
<b>Transcript N50</b>	952	1125	1277	952
<b>Median transcript length</b>	416	455	446	417
<b>Mean transcript length</b>	663.57	732.7	777.18	663.29

Table 5.2. Assembly statistics after successive filtering by IsoPct and FPKM, Swiss-prot annotations, and Decon-Seq.

	<i>I. batatas</i>	<i>I. trifida</i>	<i>M. dissecta</i>	<i>M. quinquefolia</i>
<b>Transcripts in original assembly</b>	245140	119153	254174	363820
<b>Transcripts filtered by IsoPct, FPKM</b>	158267 (64.6%)	51181 (43.0%)	176584 (69.5%)	209593 (57.6%)
<b>Transcripts filtered by Swiss-prot annotations</b>	5097 (2.1%)	5262 (4.4%)	3529 (1.4%)	3665 (1.0%)
<b>Transcripts filtered by Decon-Seq</b>	619 (0.3%)	491 (0.4%)	441 (0.2%)	595 (0.2%)
<b>Total removed by filtering</b>	163983 (66.9%)	56934 (47.8%)	180554 (71.0%)	213853 (58.8%)

Table 5.3. Top ten most abundant gene ontology (GO) categories represented in genes differentially expressed between storage roots and fine roots of *Ipomoea batatas* and *Merremia dissecta* considering each species separately.

<b>Species</b>	<b>% of Total</b>	<b>GO annotation</b>	<b>type</b>	<b>specific</b>
<i>I. batatas</i>	4.50	GO:0016021	cellular_component	integral component of membrane
<i>I. batatas</i>	3.12	GO:0005634	cellular_component	nucleus
<i>I. batatas</i>	2.96	GO:0005886	cellular_component	plasma membrane
<i>I. batatas</i>	2.45	GO:0005524	molecular_function	ATP binding
<i>I. batatas</i>	1.98	GO:0046872	molecular_function	metal ion binding
<i>I. batatas</i>	1.89	GO:0006351	biological_process	transcription, DNA-templated
<i>I. batatas</i>	1.69	GO:0005576	cellular_component	extracellular region
<i>I. batatas</i>	1.68	GO:0009507	cellular_component	chloroplast
<i>I. batatas</i>	1.63	GO:0003700	molecular_function	sequence-specific DNA binding transcription factor activity
<i>I. batatas</i>	1.57	GO:0003677	molecular_function	DNA binding
<i>M. dissecta</i>	4.37	GO:0016021	cellular_component	integral component of membrane
<i>M. dissecta</i>	3.06	GO:0005634	cellular_component	nucleus
<i>M. dissecta</i>	2.51	GO:0005886	cellular_component	plasma membrane
<i>M. dissecta</i>	2.51	GO:0003700	molecular_function	sequence-specific DNA binding transcription factor activity
<i>M. dissecta</i>	2.18	GO:0006351	biological_process	transcription, DNA-templated
<i>M. dissecta</i>	2.07	GO:0005524	molecular_function	ATP binding
<i>M. dissecta</i>	1.86	GO:0009507	cellular_component	chloroplast
<i>M. dissecta</i>	1.53	GO:0006355	biological_process	regulation of transcription, DNA-templated
<i>M. dissecta</i>	1.53	GO:0003677	molecular_function	DNA binding
<i>M. dissecta</i>	1.42	GO:0009501	cellular_component	amyloplast

Table 5.4. Top ten most abundant gene ontology (GO) categories represented in the set of orthologous genes differentially expressed between storage roots and fine roots in both *Ipomoea batatas* and *Merremia dissecta*.

Species	% of Total	GO annotation	type	Specific
<i>I. batatas</i>	3.89	GO:0009507	cellular_component	chloroplast
<i>I. batatas</i>	3.53	GO:0016021	cellular_component	integral component of membrane
<i>I. batatas</i>	3.18	GO:0009501	cellular_component	amyloplast
<i>I. batatas</i>	3.18	GO:0005634	cellular_component	Nucleus
<i>I. batatas</i>	2.83	GO:0005524	molecular_function	ATP binding
<i>I. batatas</i>	2.12	GO:0019252	biological_process	starch biosynthetic process
<i>I. batatas</i>	2.12	GO:0003700	molecular_function	sequence-specific DNA binding transcription factor activity
<i>I. batatas</i>	1.77	GO:0006351	biological_process	transcription, DNA-templated
<i>I. batatas</i>	1.77	GO:0005886	cellular_component	plasma membrane
<i>I. batatas</i>	1.77	GO:0003677	molecular_function	DNA binding
<i>M. dissecta</i>	4.00	GO:0009507	cellular_component	chloroplast
<i>M. dissecta</i>	4.00	GO:0016021	cellular_component	integral component of membrane
<i>M. dissecta</i>	3.20	GO:0009501	cellular_component	amyloplast
<i>M. dissecta</i>	3.20	GO:0005634	cellular_component	Nucleus
<i>M. dissecta</i>	2.40	GO:0005576	cellular_component	extracellular region
<i>M. dissecta</i>	2.40	GO:0005524	molecular_function	ATP binding
<i>M. dissecta</i>	2.40	GO:0003677	molecular_function	DNA binding
<i>M. dissecta</i>	2.00	GO:0019252	biological_process	starch biosynthetic process
<i>M. dissecta</i>	2.00	GO:0006351	biological_process	transcription, DNA-templated
<i>M. dissecta</i>	2.00	GO:0003700	molecular_function	sequence-specific DNA binding transcription factor activity

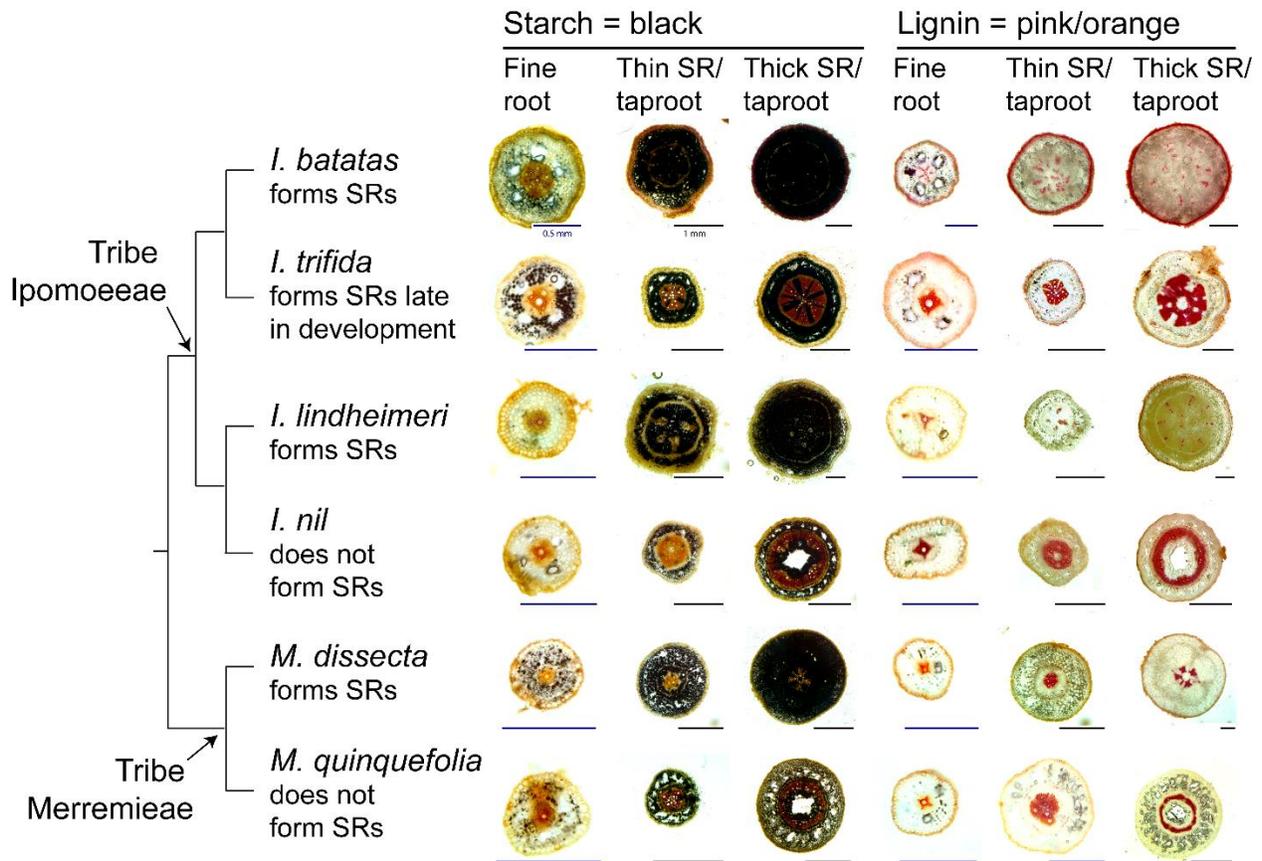


Figure 5.1. Root cross sections from three pairs of species, where one member of the species pair forms storage roots and the other does not. To the left is a phylogeny depicting the evolutionary relationships among the six species with arrows denoting the two tribes, Ipomoeae and Merremieae. The left-most three columns are root sections stained with Lugol's iodine, which indicates starch a dark blue to black color. The right-most three columns are root sections stained with phloroglucinol-HCl, which stains lignin orange to pink. Scale bars are included with each section. Black bars are 1mm, and blue bars are 0.5 mm in length.

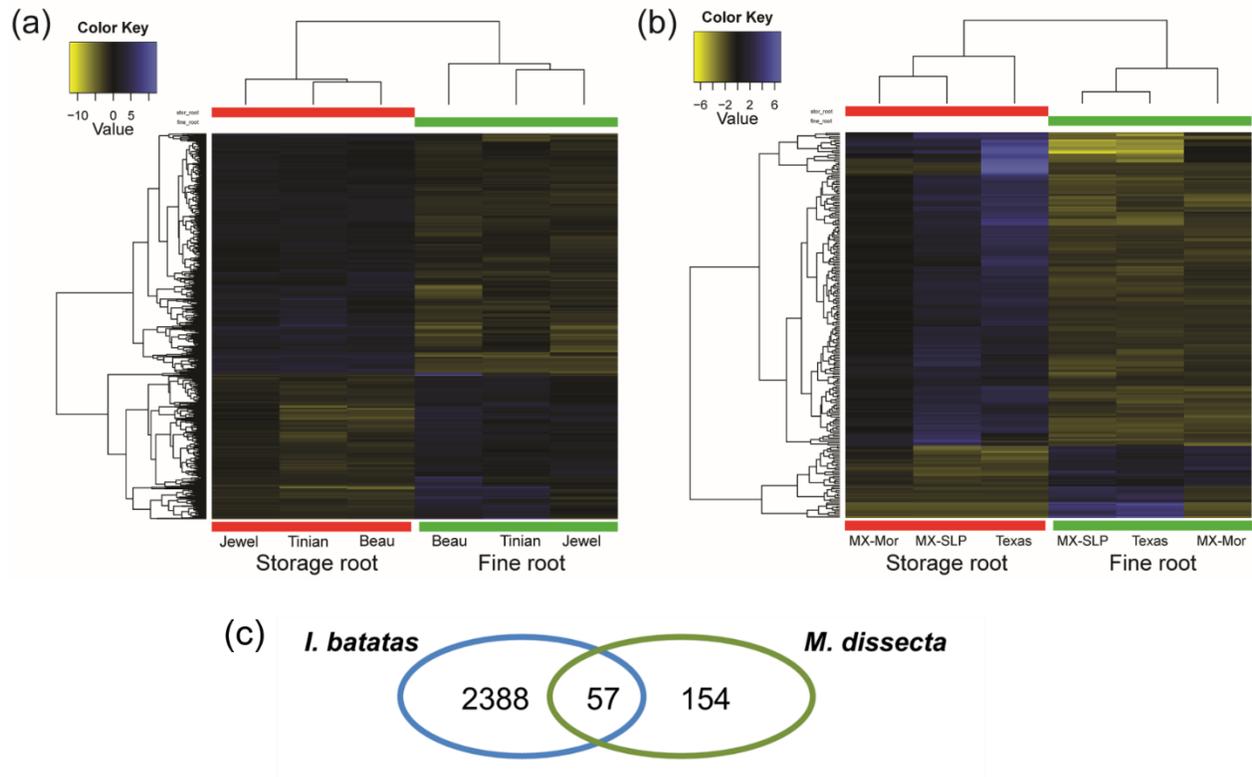


Figure 5.2. Heat map of genes differentially expressed between storage roots and fine roots of sweet potato, *Ipomoea batatas* (a) and *Merremia dissecta* (b). Each row in the heatmap is depicting the expression patterns of each transcript, and each column represents each library. A dendrogram illustrating clustering of libraries is shown above each heatmap, and a dendrogram showing clustering of transcript expression patterns is to the left of each heatmap. (c) Number of transcripts differentially expressed between storage and fine roots and the number that were orthologous between *I. batatas* and *M. dissecta*.

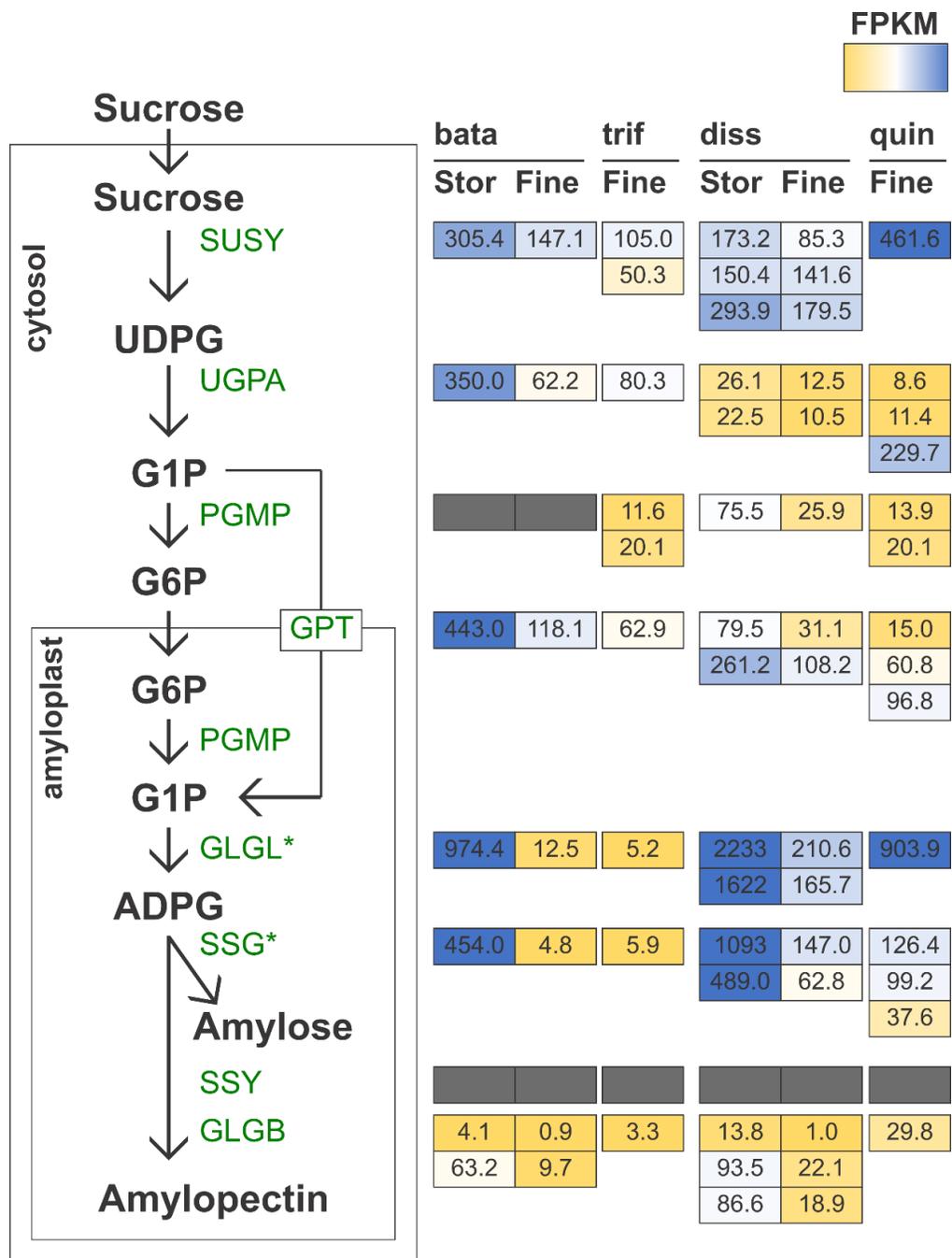


Figure 5.3. Starch biosynthetic pathway adapted from Bahaji et al. 2014. Metabolites are shown in black, and enzymes are shown in green. Shown are TMM-normalized FPKM values for homologs in all four species (bata = *Ipomoea batatas*, trif = *I. trifida*, diss = *Merremia dissecta*, and quin = *M. quinquefolia*). Grey boxes indicate genes where orthology could not be determined. Stacked boxes indicate homologs of a particular gene. Gene names with an asterisk were found to be significantly differentially expressed at a FDR < 0.05 in both *I. batatas* and *M. dissecta*. The heatmap is colored by percentile, where genes in the 10<sup>th</sup> percentile were colored yellow and those in the 90<sup>th</sup> percentile were colored dark blue.

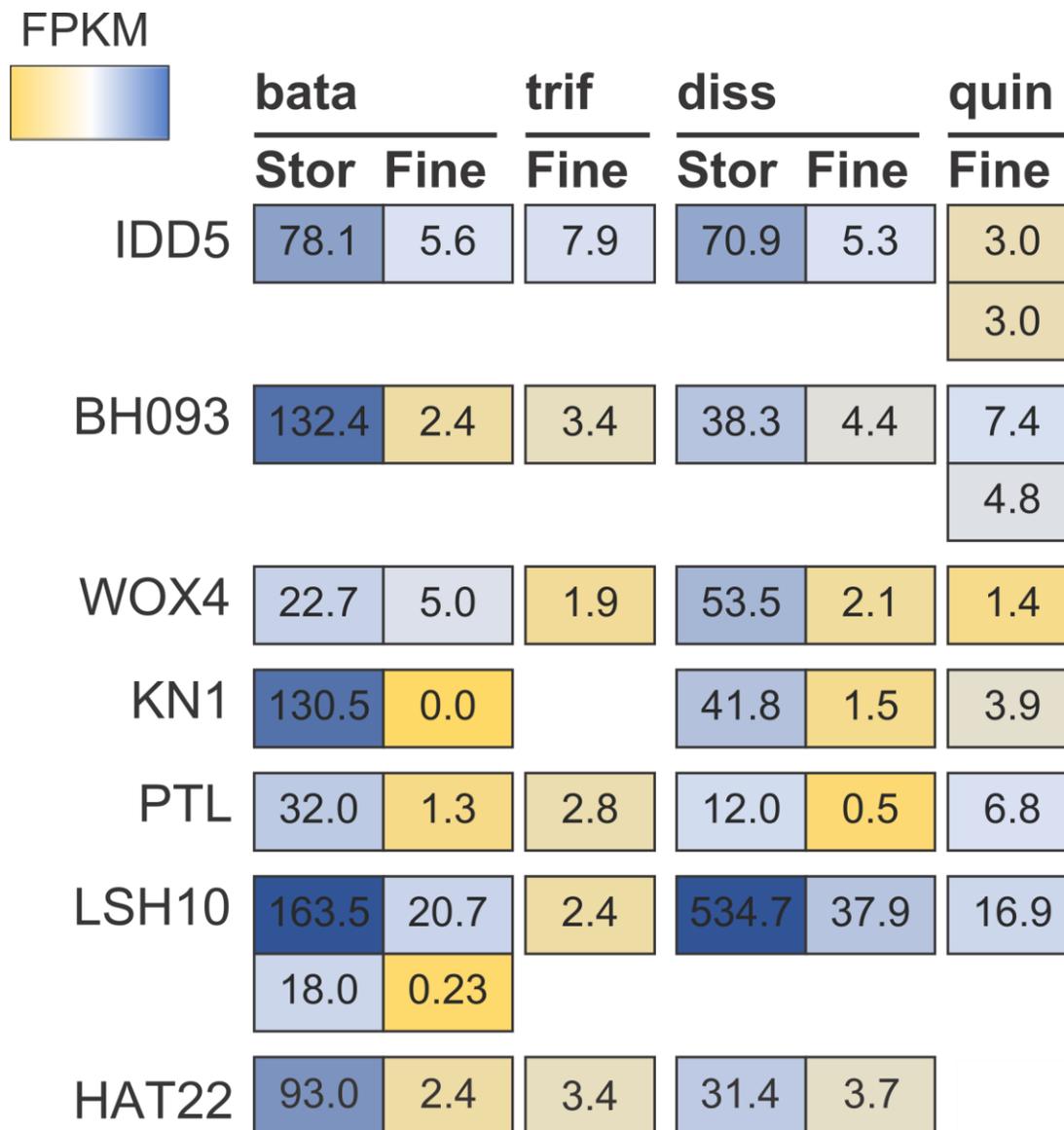


Figure 5.4. Mean TMM-normalized FPKM values for the seven transcription factors found to be significantly differentially expressed between storage and fine roots in both *Ipomoea batatas* and *Merremia dissecta* at a FDR <0.05. The heatmap depicts mean TMM-normalized FPKM values for orthologs of the transcription factors in each tissue type for all four species (bata = *Ipomoea batatas*, trif = *I. trifida*, diss = *Merremia dissecta*, and quin = *M. quinquefolia*). The heatmap was colored by percentile, where genes in the 10<sup>th</sup> percentile were colored yellow and those in the 90<sup>th</sup> percentile were colored dark blue. No ortholog of KN1 could be identified in the transcriptome assembly of *I. trifida*, and no ortholog of HAT22 could be identified in *M. quinquefolia*.

## CHAPTER VI:

### CONCLUSION AND DISCUSSION

#### **Morning glory phylogenomics**

##### *Ipomoeae*

Prior to this work, evolutionary relationships among major lineages in the Ipomoeae relied solely on a small amount of molecular data. These studies made significant advances in our understanding of relationships and primarily showed that morning glory taxonomy is incongruent with molecular results (McDonald and Mabry 1992; Miller et al. 1999, 2002, 2004; Manos et al. 2001). The results of this work largely agrees with previous molecular phylogenetic results in morning glories. Morning glory species, which based on taxonomy would be considered distantly related, were in fact closely related. Furthermore, this work adds a temporal context to morphological evolution. We showed that in the ca. 35 million years of morning glory diversification, there were no obvious fixed morphological differences between the two major morning glory lineages. Morphological evolution in morning glories is highly labile (Manos et al. 2001) and is likely driven by genetic drift, e.g. loss of association with ergot alkaloid producing fungi, or fine scale environmental pressures, e.g. storage root evolution.

##### *The Batatas complex*

Previous work estimating the phylogeny of the Batatas complex generally showed contradictory findings (Austin 1988b; Jarret et al. 1992; Rajapakse et al. 2004). This is likely due to estimating relationships using single-gene phylogenies. Using a large number of loci, we reconstructed relationships among members of the Batatas complex using a small number of taxa

(Chapter 3) and using population-level sampling (Chapter 4). Phylogenies estimated using both sets of taxa showed that incomplete lineage sorting is a major source of gene tree discordance in the Batatas complex. In addition, a large degree of discordance can be attributed to ancient hybridization followed by species diversification. Much of the literature on ancient hybridization is in reference to ancient hybrid species (e.g. (Doebley et al. 1984; Rieseberg et al. 1996, 2003; Wolfe et al. 1998)). In contrast, much less focus is given to ancient hybridization prior to speciation, likely because this pattern has been relatively difficult to discern until recently developed tools in phylogenetic network construction have become available. Given the high frequency of hybridization in plant species, it is likely that advances in phylogenomics and analyses inferring hybridization with phylogenomic data will likely reveal this pattern in more plant groups. Specifically in the Batatas complex, it is clear that ancient hybridization followed by species diversification has shaped the evolutionary history of this group.

### **Storage root evolution**

This study is among the first to investigate the evolution of storage roots. In Chapter 2, we discovered that storage root evolution is highly labile, and storage roots were derived independently at least ten times in morning glories. This was likely through parallel rather than convergent evolution (Chapter V). These results mirror a study of storage root evolution in *Adenia*, a genus of ca. 100 species in the Passifloraceae, which found that storage roots had evolved at least five times independently (Hearn 2006). Further examination of root anatomy demonstrated that these independent origins of storage roots in *Adenia* were through parallel rather than convergent evolution (Hearn 2009). Taken together, these results suggest that a storage root is a complex morphological trait that evolves independently through similar genetic mechanisms, at least in these two families of eudicots.

However, it remains unclear what drives the evolution of storage roots. There is a very weak correlation between root traits and genome size in the Batatas complex, but genome size explains very little of the variation observed in root traits. This suggests that factors other than genome size are necessary to explain the diversity of root traits in the Batatas complex. These factors are likely related to local environmental conditions. In many cases, storage roots have been cited as an example of an adaptation to harsh environmental conditions, such as fire or drought (Pate et al. 1990; Bell et al. 1996; Brunner et al. 2015). In sweetpotato, storage root bulking and growth is additionally determined by several environmental factors such as water availability (Andrade et al. 2016), temperature (Villavicencio et al. 2007), and light availability (Loretan et al. 1994). In some species, storage root size and carbohydrate content actually increases under drought conditions (Galvez et al. 2011).

In addition, the way in which storage roots are measured and described can be complicated. Storage roots are often thought of as a binary trait, where plants either have storage roots or do not store starch in roots. This is an oversimplification of a complex morphological trait. There are certainly species which inhabit the extremes of the storage root spectrum. For example, accessions of *Ipomoea lacunosa* in Chapter 4 showed very thin taproots and lateral roots and had very low starch concentration. In contrast, some closely related accessions of *I. cordatotriloba* had thicker taproots with higher starch concentration. In addition, many accessions had root widths and starch concentrations that varied greatly among accessions.

This research has provided a number of insights into the evolution and development of storage roots in morning glories, which were previously not known. Moving forward, it would be interesting to examine the effect of environmental conditions such as water limitation on storage root growth and development in a selected subset of sweetpotato wild relatives. Understanding

which factors are most relevant to storage root growth and development may provide insights into the factors which have influenced the evolution of storage roots. Furthermore, a time series study examining the growth and development of storage roots through various developmental stages would also provide insight into the early changes associated with storage root formation.

## References

- Andrade M.I., Naico A., Ricardo J., Eyzaguirre R., Makunde G.S., Ortiz R., Grüneberg W.J. 2016. Genotype × environment interaction and selection for drought adaptation in sweetpotato (*Ipomoea batatas* [L.] Lam.) in Mozambique. *Euphytica*. 209:261–280.
- Austin D.F. 1988. The taxonomy evolution and genetic diversity of sweetpotatoes and related wild species. In: Gregory P., editor. *Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources: Report of the 1st sweet potato planning conference 1987*. Lima, Peru: CIP. p. 27–59.
- Bell T.L., Pate J.S., Dixon K.W. 1996. Relationships Between Fire Response, Morphology, Root Anatomy and Starch Distribution in South-west Australian Epacridaceae. *Ann. Bot.* 77:357–364.
- Brunner I., Herzog C., Dawes M.A., Arend M., Sperisen C. 2015. How tree roots respond to drought. *Front. Plant Sci.* 6:1–16.
- Doebley J., Goodman M.M., Stuber C.W. 1984. Isoenzymatic variation in *Zea* (Gramineae). *Syst. Bot.* 9:203–218.
- Galvez D.A., Landhäusser S.M., Tyree M.T. 2011. Root carbon reserve dynamics in aspen seedlings: Does simulated drought induce reserve limitation? *Tree Physiol.* 31:250–257.
- Hearn D.J. 2006. *Adenia* (Passifloraceae) and its adaptive radiation: Phylogeny and growth form diversification. *Syst. Bot.* 31:805–821.
- Hearn D.J. 2009. Developmental patterns in anatomy are shared among separate evolutionary origins of stem succulent and storage root-bearing growth habits in *Adenia* (Passifloraceae). *Am. J. Bot.* 96:1941–56.
- Jarret R.L., Gawel N., Whittemore A.T. 1992. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* L.] Lam.]. *J. Am. Soc. Hortic. Sci.* 117:633–637.
- Loretan P.A., Bonsi C.K., Mortley D.G., Wheeler R.M., Mackowiak C.L., Hill W.A., Morris C.E., Trotman A.A., David P.P. 1994. Effects of several environmental factors on sweetpotato growth. *Adv. Sp. Res.* 14:277–80.

- Manos P.S., Miller R.E., Wilkin P. 2001. Phylogenetic Analysis of *Ipomoea*, *Argyreia*, *Stictocardia*, and *Turbina* Suggests a Generalized Model of Morphological Evolution in Morning Glories. *Syst. Bot.* 26:585–602.
- McDonald J.A., Mabry T.J. 1992. Phylogenetic systematics of New World Convolvulaceae based on chloroplast DNA restriction site variation. *Plant Syst. Evol.* 180:243–259.
- Miller R.E., Buckley T.R., Manos P.S. 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Syst. Biol.* 51:740–753.
- Miller R.E., McDonald J.A., Manos P.S. 2004. Systematics of *Ipomoea* subgenus *Quamoclit* based on ITS sequence data and a Bayesian phylogenetic analysis. *Am. J. Bot.* 91:1208–1218.
- Miller R.E., Rausher M.D., Manos P.S. 1999. Phylogenetic Systematics of *Ipomoea* (Convolvulaceae) Based on ITS and Waxy Sequences. *Syst. Bot.* 24:209–227.
- Pate J.S., Froend R.H., Bowen B.J., Hansen A., Kuo J. 1990. Seedling growth and storage characteristics of seeder and resprouter species of Mediterranean-type ecosystems of SW Australia. *Ann. Bot.* 65:585–601.
- Rajapakse S., Nilmalgoda S.D., Molnar M., Ballard R.E., Austin D.F., Bohac J.R. 2004. Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear beta-amylase gene sequences. *Mol. Phylogenet. Evol.* 30:623–32.
- Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., Lexer C. 2003. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science.* 301:1211–1216.
- Rieseberg L.H., Sinervo B., Linder C.R., Ungerer M.C., Arias D.M. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science.* 272:741–745.
- Villavicencio L.E., Blankenship S.M., Yencho G.C., Thomas J.F., Raper C.D. 2007. Temperature Effect on Skin Adhesion, Cell Wall Enzyme Activity, Lignin Content, Anthocyanins, Growth Parameters, and Periderm Histochemistry of Sweetpotato. *J. Am. Soc. Hortic. Sci.* 132:729–738.
- Wolfe A.D., Xiang Q.-Y., Kephart S.R. 1998. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proc. Natl. Acad. Sci.* 95:5112–5115.

## APPENDIX A

## SUPPLEMENTARY FIGURES AND TABLES – CHAPTER II

Table S2.1 – Voucher, GenBank accession, and locality information for species included in this study.

<b>Taxon</b>	<b>Accession number</b>	<b>plastome GenBank accession number</b>	<b>Voucher specimen, Collection locale, Herbarium</b>
<i>Argyreia nervosa</i> (Burm. f.) Bojer	REM 77	KF242477	B&T World Seeds 52121, SELU.
<i>Ipomoea amnicola</i> Morong	REM 36	KF242478	G. Lowe, PI 553010, United States, Texas, SELU.
<i>I. argillicola</i> R. W. Johnson	REM 38	KF242479	R. Jarret 7531a, Australia, Queensland, SELU.
<i>I. batatas</i> (L.) Lam.	PI 508520	KF242473	PI 508520, China, maintained in vitro at USDA Plant Genetic Resources Unit, Griffin, Georgia.
<i>I. batatas</i> (L.) Lam.	PI 518474	KF242474	D. Austin & F. de La Puente, Mexico, Veracruz, GA.
<i>I. batatas</i> (L.) Lam.	PI 561258	KF242475	D. Austin, Ecuador, El Oro, GA.
<i>I. cairica</i> (L.) Sweet	REM 184	KF242480	B. Burson & W. Langford 559, Uruguay, SELU.
<i>I. cordatotriloba</i> D. F. Austin	REM 317	KF242497	R. Miller 317, United States, Louisiana, SELU.
<i>I. diamantinensis</i> J. W. Black	REM 37	KF242481	R. Johnson J281, Australia, Queensland, SELU.
<i>I. dumetorum</i> Willd. ex Roem & Schult	REM 218	KF242482	J. A. McDonald 140, United States, New Mexico, SELU.
<i>I. eriocarpa</i> R. Br.	REM 190	KF242483	R. Johnson J50, Australia, Queensland, SELU.
<i>I. hederifolia</i> L.	REM 476	KF242484	R. Miller MX05-42, Mexico, Oaxaca, SELU.
<i>I. involucrata</i> F. Dietr. ex Choisy	REM 851	KF242485	Kew 212975, Mali, Segou, SELU.
<i>I. minutiflora</i> (M. Martens & Galeotti) House	REM 535	KF242498	M. T. Clegg 56, SELU.
<i>I. murucoides</i> Roem. & Schult.	REM 351	KF242486	R. Miller MX04-03, Mexico, Michoacán, SELU.
<i>I. nil</i> (L.) Roth	REM 459	KF242487	R. Miller MX05-52, Mexico, Oaxaca, SELU.
<i>I. obscura</i> (L.) Ker Gawl.	REM 271	KF242499	T. C. Mendelson s.n., United States, Hawaii, SELU.
<i>I. orizabensis</i> (H. B. K.) G. Don	REM 178	KF242488	M. Rausher s.n., Mexico, SELU.
<i>I. pedicellaris</i> Benth.	REM 403	KF242489	R. Miller MX04-39, Mexico, Jalisco, SELU.

<i>I. pes-caprae</i> (L.) R. Br.	REM 767	KF242490	R. Miller CR08-21, Costa Rica, Puntarenas, SELU.
<i>I. pes-tigridis</i> L.	REM 854	KF242500	Kew 572, India, Uttar Pradesh, SELU.
<i>I. polpha</i> R. W. Johnson	REM 85	KF242491	SBE Universal Seed Bank, SELU.
<i>I. setosa</i> Ker. Gawl	REM 68	KF242492	Collector unknown, United States, Texas, SELU.
<i>I. splendor-sylvae</i> House	REM 763	KF242493	R. Miller CR08-17, Costa Rica, Puntarenas, SELU.
<i>I. ternifolia</i> Cav.	REM 452	KF242494	R. Miller MX05-41, Mexico, Oaxaca, SELU.
<i>I. tricolor</i> Cav.	REM 448	KF242495	R. Miller MX05-33, Mexico, Oaxaca, SELU.
<i>I. trifida</i> G. Don	REM 753	KF242496	R. Miller CR08-04, Costa Rica, Guanacaste, SELU.
<i>I. trifida</i> G. Don	PI 618966	KF242476	PI 618966, Mexico, Michoacán, GA.
<i>Merremia quinquefolia</i> Hallier f.	REM 389	KF242501	R. Miller MX04-27, Mexico, Jalisco, SELU.
<i>Operculina macrocarpa</i> Urb.	REM 205	KF242502	Kew 98108, SELU.
<i>Stictocardia macalusoii</i> (Mattei) Verdc.	REM 206	KF242503	Kew 97536, Oman, Dhofar, SELU.
<i>Turbina corymbosa</i> (L.) Raf.	REM 855	KF242504	K. Clay, SELU.

Table S2.2 – Species used in this analysis, accession number, sequencing technology used for a particular accession, number of raw reads, and chloroplast genome coverage.

<b>Taxon</b>	<b>Accession</b>	<b>Sequencing technology</b>	<b>No. reads</b>	<b>Depth of Coverage</b>
<i>Argyreia nervosa</i> (Burm. f.) Bojer	REM 77	454	22534	12
<i>I. amnicola</i> Morong	REM 36	Illumina	804616	16
<i>I. argillicola</i> R.W. Johnson	REM 38	Illumina	1912585	139
<i>I. batatas</i> (L.) Lam.	PI 508520	Illumina	67271204	3498
<i>I. batatas</i> (L.) Lam.	PI 518474	Illumina	70753160	5524
<i>I. batatas</i> (L.) Lam.	PI 561258	Illumina	93729744	4443
<i>I. cairica</i> (L.) Sweet	REM 184	Illumina	1159645	35
<i>I. cordatotriloba</i> Dennst.	REM 317	Illumina	3182397	139
<i>I. diamantinensis</i> J.M. Black	REM 37	Illumina	4164154	359
<i>I. dumetorum</i> Willd. ex Roem. & Schult.	REM 218	Illumina	3597756	232
<i>I. eriocarpa</i> R. Br.	REM 190	Illumina	3809511	174
<i>I. hederifolia</i> L.	REM 476	Illumina	4690471	218
<i>I. involucrata</i> P. Beauv.	REM 851	Illumina	4350605	174
<i>I. minutiflora</i> (M. Martens & Galeotti) House	REM 535	Illumina	1690482	58
<i>I. murucoides</i> Roem. & Schult.	REM 351	Illumina	248626	8
<i>I. nil</i> (L.) Roth	REM 459	Illumina	2405014	131
<i>I. obscura</i> (L.) Ker Gawl.	REM 271	Illumina	5786721	216
<i>I. orizabensis</i> (G. Pelletan) Ledeb. ex Steud.	REM 178	Illumina	681853	29
<i>I. pedicellaris</i> Benth.	REM 403	Illumina	674895	29
<i>I. pes-caprae</i> (L.) R. Br.	REM 767	454	11706	11
<i>I. pes-tigridis</i> L.	REM 854	Illumina	4632352	100
<i>I. polpha</i> R.W. Johnson	REM 85	Illumina	7986552	169
<i>I. setosa</i> Ker Gawl.	REM 68	Illumina	4107725	123
<i>I. splendor-sylvae</i> House	REM 763	Illumina	2582382	142
<i>I. ternifolia</i> Cav.	REM 452	Illumina	3606511	257
<i>I. tricolor</i> Cav.	REM 448	454	34003	25
<i>I. trifida</i> (Kunth) G. Don	REM 753	Illumina	3529757	230
<i>I. trifida</i> (Kunth) G. Don	PI 618966	Illumina	98134098	7261
<i>Stictocardia macalusoii</i> (Mattei) Verdc.	REM 206	Illumina	737018	26
<i>Turbina corymbosa</i> (L.) Raf.	REM 855	Illumina	1122206	44
<i>Merremia quinquefolia</i> (L.) Hallier f.	REM 389	Illumina	868647	39
<i>Operculina macrocarpa</i> (L.) Urb.	REM 205	Illumina	6473772	227

Table S2.3 – Species used in this analysis with GC content and placement of inverted repeat boundaries.

Species	Accession	% GC	LSC-IR <sub>A</sub>	IR <sub>A</sub> -SSC	SSC-IR <sub>B</sub>	IR <sub>B</sub> -LSC
<i>Argyrea nervosa</i> (Burm. f.) Bojer	REM 77	37	ycf2	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. amnicola</i> Morong	REM 36	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. argillicola</i> R.W. Johnson	REM 38	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. batatas</i> (L.) Lam.	PI 508520	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. batatas</i> (L.) Lam.	PI 518474	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. batatas</i> (L.) Lam.	PI 561258	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. cairica</i> (L.) Sweet	REM 184	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. cordatotriloba</i> Dennst.	REM 317	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. diamantinensis</i> J.M. Black	REM 37	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. dumetorum</i> Willd. ex Roem. & Schult.	REM 218	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. eriocarpa</i> R. Br.	REM 190	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. hederifolia</i> L.	REM 476	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. involucrata</i> P. Beauv.	REM 851	37	ycf2	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. minutiflora</i> (M. Martens & Galeotti) House	REM 535	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. murucoides</i> Roem. & Schult.	REM 351	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. nil</i> (L.) Roth	REM 459	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. obscura</i> (L.) Ker Gawl.	REM 271	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. orizabensis</i> (G. Pelletan) Ledeb. ex Steud.	REM 178	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. pedicellaris</i> Benth.	REM 403	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. pes-caprae</i> (L.) R. Br.	REM 767	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. pes-tigridis</i> L.	REM 854	37	rpl23-trnI	ndhH-ndhF	ndhA-intron	trnH-trnI
<i>I. polpha</i> R.W. Johnson	REM 85	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. setosa</i> Ker Gawl.	REM 68	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. splendor-sylvae</i> House	REM 763	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. ternifolia</i> Cav.	REM 452	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. tricolor</i> Cav.	REM 448	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. trifida</i> (Kunth) G. Don	REM 753	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>I. trifida</i> (Kunth) G. Don	PI 618966	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>Stictocardia macalusoi</i> (Mattei) Verdc.	REM 206	37	trnI-ycf2	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>Turbina corymbosa</i> (L.) Raf.	REM 855	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>Merremia quinquefolia</i> (L.) Hallier f.	REM 389	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI
<i>Operculina macrocarpa</i> (L.) Urb.	REM 205	37	rpl23-trnI	ndhH-ndhF	ndhA-exon1	trnH-trnI

Table S2.4 – Table of characters, character states, and references of each character state used in the likelihood-based ancestral character state reconstructions.

<b>Taxon</b>	<b>Root architecture</b>	<b>Root architecture reference</b>	<b>Flower color</b>	<b>Flower color reference</b>	<b>Ergot alkaloid presence</b>	<b>Ergot alkaloid reference</b>
<i>Argyreia nervosa</i> (Burm. f.) Bojer	Unknown		Purple	(Woodson et al. 1975)	Positive	(Eich 2008)
<i>I. amnicola</i> Morong	Fibrous	(O'Donell 1959)	Purple	(O'Donell 1959)	Positive	(Eich 2008)
<i>I. argillicola</i> R.W. Johnson	Tuberous	(Johnson 1986)	Purple	(Johnson 1986)	Positive	(Eich 2008)
<i>I. batatas</i> (L.) Lam.	Tuberous	(O'Donell 1959; McDonald 1994)	Purple	(McDonald 1994; Hammel 2010)	Negative	(Eich 2008)
<i>I. cairica</i> (L.) Sweet	Tuberous	(van Ooststroom and Leyden 1953)	Purple	(O'Donell 1959)	Unknown	
<i>I. cordatotriloba</i> Dennst.	Fibrous	Pers. obs.	Purple	(Austin, 1978 as <i>I. trichocarpa</i> Ell.; Sundell et al., 2002)	Unknown	
<i>I. diamantinensis</i> J.M. Black	Unknown		White	(Johnson 1992)	Positive	(Eich 2008)
<i>I. dumetorum</i> Willd. ex Roem. & Schult.	Tuberous	(Austin 1997)	Purple	(O'Donell 1959; Hammel 2010)	Positive	(Eich 2008)
<i>I. eriocarpa</i> R. Br.	Unknown		Purple	(van Ooststroom and Leyden 1953; Rhui-cheng and Staples 1995)	Negative	(Eich 2008)
<i>I. hederifolia</i> L.	Fibrous	(McDonald 1994)	Red	(McDonald 1994; Hammel 2010)	Negative	(Eich 2008)
<i>I. involucrata</i> P. Beauv.	Fibrous	Pers. obs.	Purple	(Verdcourt 1963)	Negative	(Eich 2008)
<i>I. minutiflora</i> (M. Martens & Galeotti) House	Fibrous	(McDonald 1994)	Yellow	(McDonald 1994; Hammel 2010)	Positive	(Eich 2008)
<i>I. murucoides</i> Roem. & Schult.	Not tuberous	(Austin 1997)	White	(Standley et al. 1970; McPherson 1981)	Negative	(Eich 2008)
<i>I. nil</i> (L.) Roth	Fibrous	Pers. obs.	Blue; Purple	(O'Donell 1959; McDonald 1994; Hammel 2010)	Unknown	
<i>I. obscura</i> (L.) Ker Gawl.	Fibrous	Pers. obs.	White; Yellow	(van Ooststroom and Leyden 1953; Rhui-cheng and Staples 1995)	Negative	(Eich 2008)
<i>I. orizabensis</i> (G. Pelletan) Ledeb. ex Steud.	Tuberous	(Noda et al. 1987; Linajes et al. 1994; McDonald 1994)	Purple	(House 1908; McDonald 1994)	Positive	(Eich 2008)
<i>I. pedicellaris</i> Benth.	Tuberous	(McDonald 1994)	Purple	(McDonald 1994)	Positive	(Eich 2008)
<i>I. pes-caprae</i> (L.) R. Br.	Fibrous	(McDonald 1994)	Purple	(McDonald 1994; Hammel 2010)	Positive	(Eich 2008)

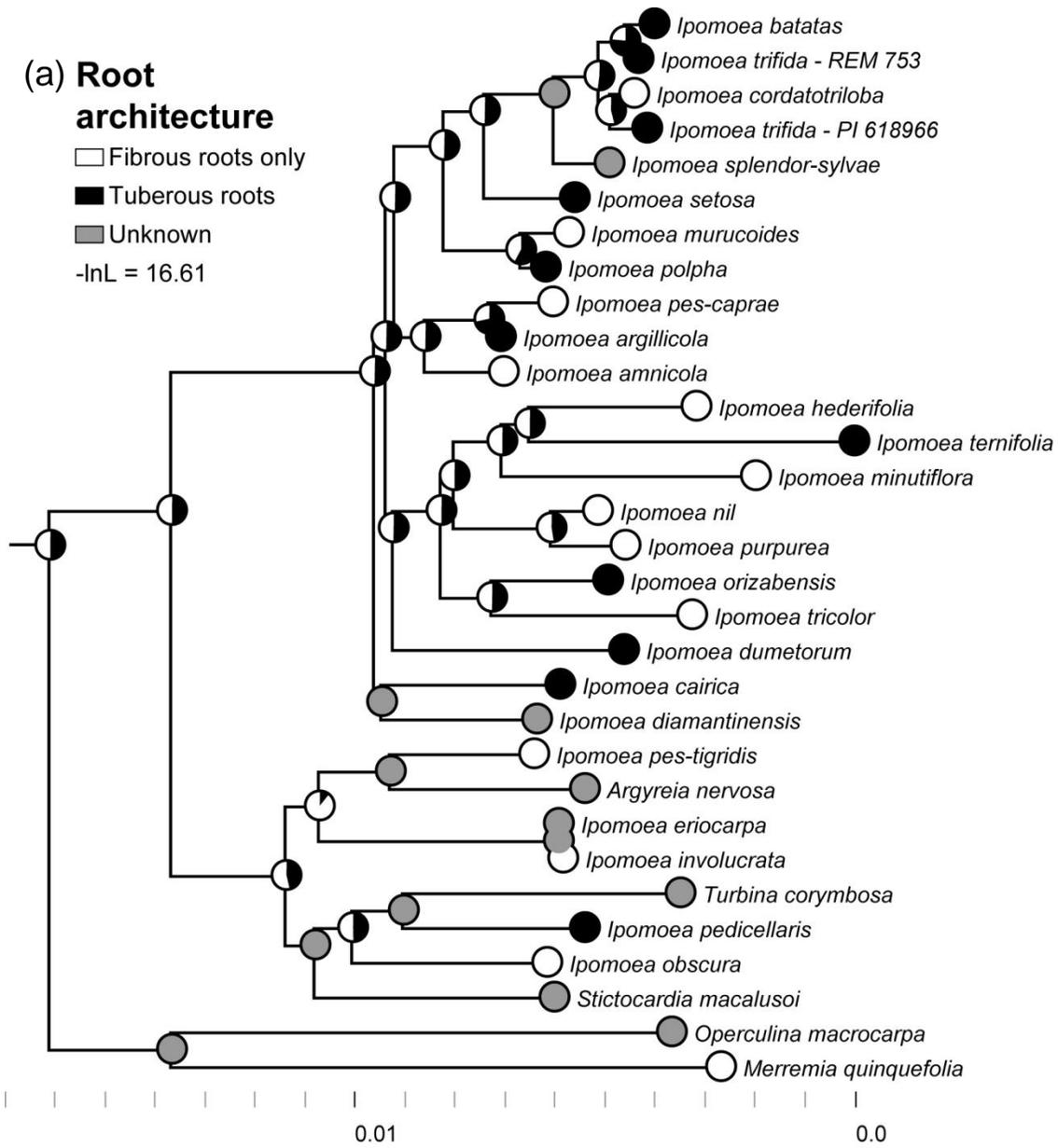
## References for Table S2.4

- Austin, D.F. 1997. Dissolution of *Ipomoea* series *Anisomerae* (Convolvulaceae). *Journal of the Torrey Botanical Society* 124: 140–159.
- Austin, D.F. 1978. The *Ipomoea batatas* complex. I. Taxonomy. *Bulletin of the Torrey Botanical Club* 105: 114–129.
- Austin, D.F., and S. Demissew. 1997. Unique fruits and generic status of *Stictocardia* (Convolvulaceae). *Kew Bulletin* 52: 161–169.
- Austin, D.F., and G.W. Staples. 1991. A revision of the neotropical species of *Turbina* Raf. (Convolvulaceae). *Bulletin of the Torrey Botanical Club* 118: 265–280.
- Eich, E. 2008. Solanaceae and Convolvulaceae Secondary Metabolites: Biosynthesis, Chemotaxonomy, Biological and Economic Significance: a handbook. Springer, Berlin.
- Hammel, B.E. 2010. Manual de plantas de Costa Rica. Vol. V. Dicotiledoneas (Clusiaceae-Gunneraceae). *Monographs in Systematic Botany from the Missouri Botanical Garden* 119: 1–970.
- House, H.D. 1908. The North American species of the genus *Ipomoea*. *Annals of the New York Academy of Sciences* 18: 181–263.
- Johnson, R.W. 1992. Convolvulaceae. In G. J. Harde [ed.], *Flora of New South Wales*, New South Wales University Press, Kensington, NSW, Australia.
- Johnson, R.W. 1986. Four new species of *Ipomoea* L. (Convolvulaceae) from Australia. *Austrobaileya* 2: 217–223.
- Linajes, A., V. Rico-Gray, and G. Carrion. 1994. Traditional production system of the root of Jalapa, *Ipomoea purga* (Convolvulaceae), in Central Veracruz, Mexico. *Economic Botany* 48: 84–89.
- McDonald, J.A. 1994. Convolvulaceae II. In V. Sosa [ed.], *Flora de Veracruz*, 1–133.
- McDonald, J.A. 1995. Revision of *Ipomoea* section *Leptocallis* (Convolvulaceae). *Harvard Papers in Botany* 6: 97–122.
- McPherson, G. 1981. Studies in *Ipomoea* (Convolvulaceae) I. The Arborescens Group. *Annals of the Missouri Botanical Garden* 68: 527–545.
- Noda, N., M. Ono, K. Miyahara, and T. Kawasaki. 1987. Resin glycosides I. Isolation and structure elucidation of orizabin-I, II, III, and IV. Genuine resin glycosides from the root of *Ipomoea orizabensis*. *Tetrahedron* 43: 3889–3902.
- O'Donnell, C.A. 1959. Convolvulaceas Argentinas. *Lilloa* 29: 87–348.

- Van Ooststroom, S.J., and Leyden. 1953. Convolvulaceae. In C. van Steenis [ed.], *Flora Malesiana*, 459–489.
- Rhui-cheng, F., and G. Staples. 1995. Convolvulaceae. In *Flora of China*, 271–325.
- Standley, P.C., L.O. Williams, and D.N. Gibson. 1970. *Flora of Guatemala, Part IX. Fieldiana: Botany* 24: 1–236.
- Sundell, E., R.D. Thomas, C. Amason, and C. Doffitt. 2002. Noteworthy vascular plants from Arkansas. II. *SIDA* 20: 409–418.
- Urban, I. 1903. *Nova genera et species II*. In I. Urban [ed.], *Synbolae Antillanae: seu Fundamenta Florae Indiae Occidentalis*, Paul Klincksieck, Paris, London.
- Verdcourt, B. 1963. Convolvulaceae. In C. E. Hubbard, and E. Milne-Redhead [eds.], *Flora of tropical East Africa*, 1–161. Whitefriars Press, London.
- Woodson, R.E., R.W. Schery, and D.F. Austin. 1975. *Flora of Panama. Part IX. Family 164. Convolvulaceae*. *Annals of the Missouri Botanical Garden* 62: 157–224.

(a) **Root architecture**

- Fibrous roots only
  - Tuberous roots
  - Unknown
- lnL = 16.61



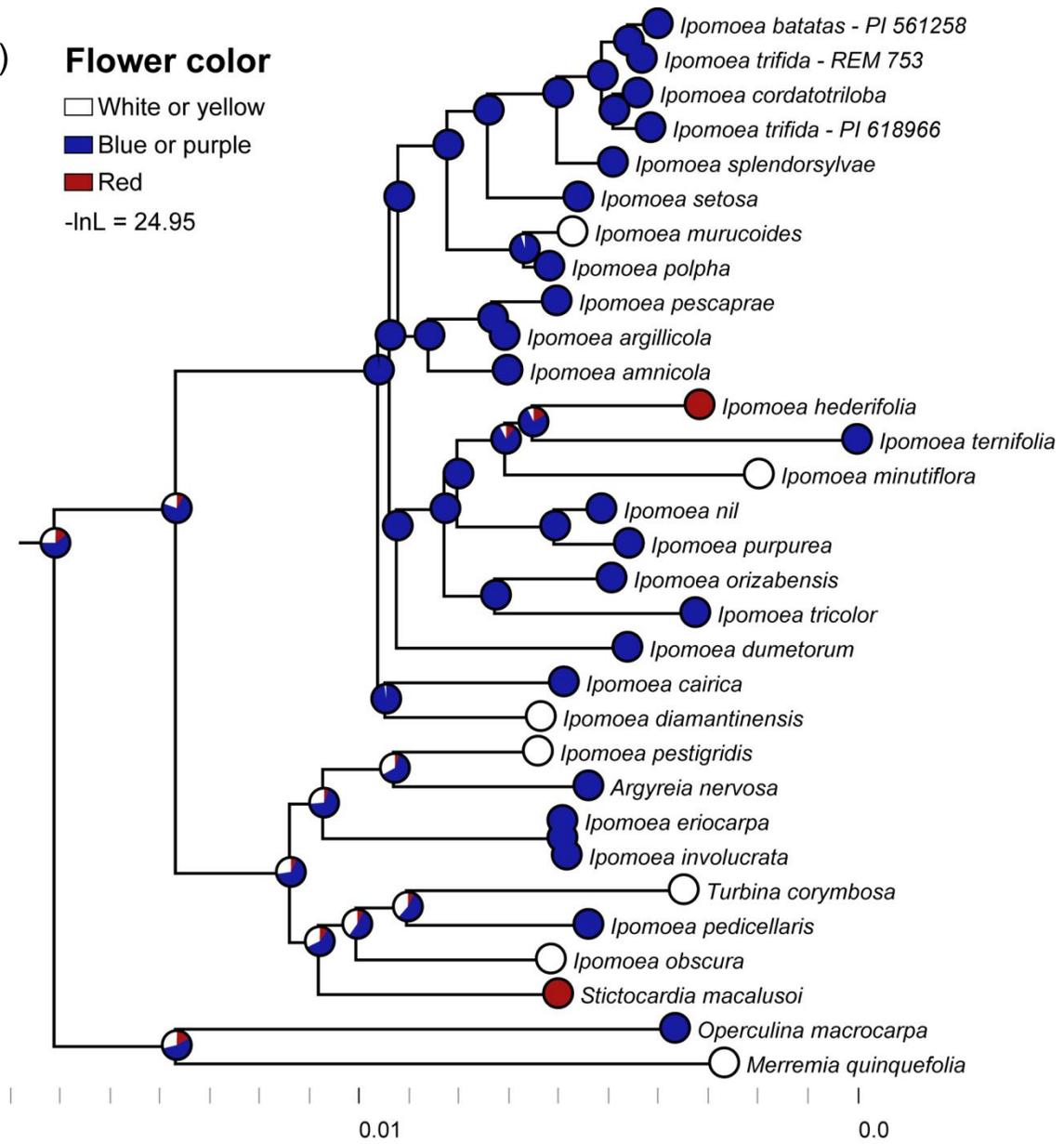
(b) **Flower color**

□ White or yellow

■ Blue or purple

■ Red

-lnL = 24.95





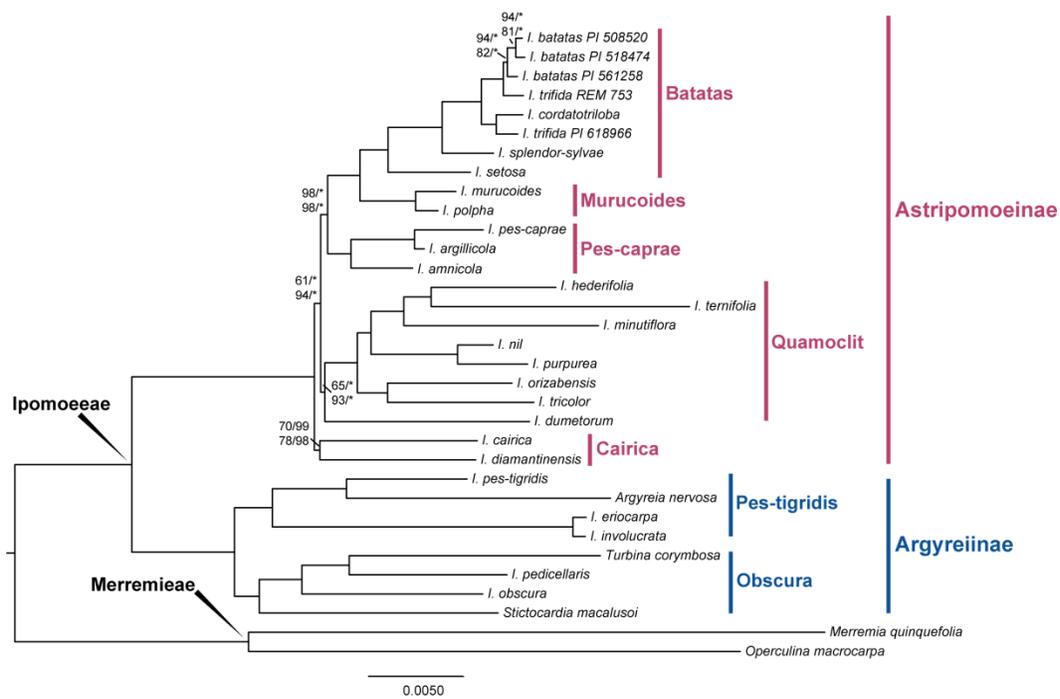


Figure S2.2 – Phylogeny of the Ipomoeae based on 82 protein coding and rRNA sequences. Genes from the second inverted repeat region were removed for analyses. The topology shown is from a maximum likelihood analysis in RAxML of the MUSCLE alignment. Nodes without numbers or with an asterisk (\*) received 100% BS and PP support in all analyses. Numbers behind nodes are maximum likelihood bootstrap (BS) and Bayesian posterior probability (PP) values for the MUSCLE and SATé alignments. Top numbers are Muscle BS and PP values. Lower numbers are SATé BS and PP values. Pink bars to the right are well-supported lineages within the Astripomoeinae; blue bars are lineages within the Argyreiinae.

APPENDIX B

SUPPLEMENTARY FIGURES AND TABLES – CHAPTER IV

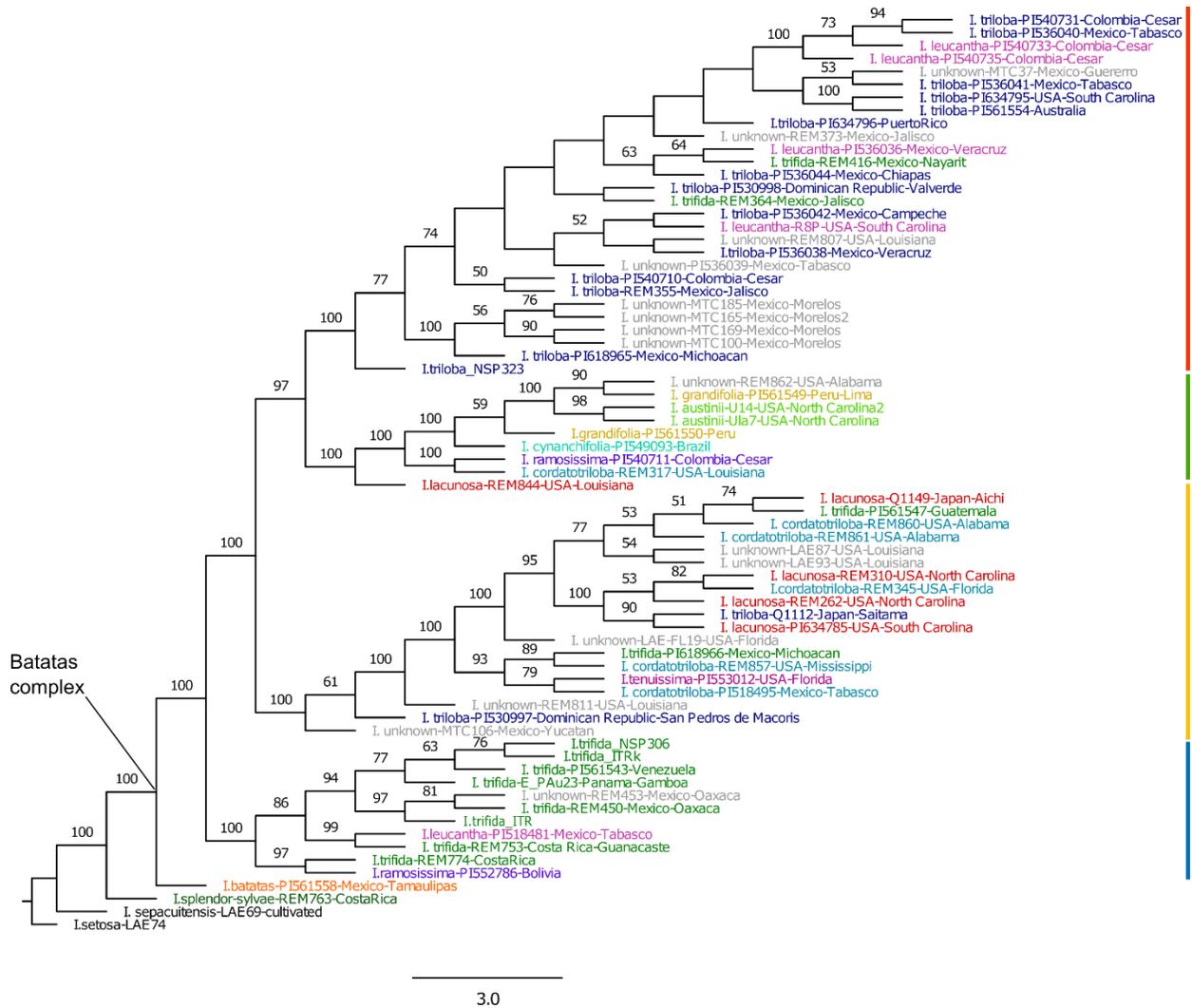


Figure S4.1 – Phylogeny of the Batatas complex estimated in SVDQuartets including only diploid individuals. Numbers behind nodes indicate bootstrap support from bootstrapping done in SVDQuartets. Nodes without numbers had bootstrap support <50%. Each individual is colored by species designation. Grey individuals are those which exhibited intermediate traits and could not confidently be identified to species using existing taxonomic keys. Colored bars to the right of the phylogeny indicate well-supported clades recovered in the concatenation, ASTRAL-II, and SVDQuartets analyses.



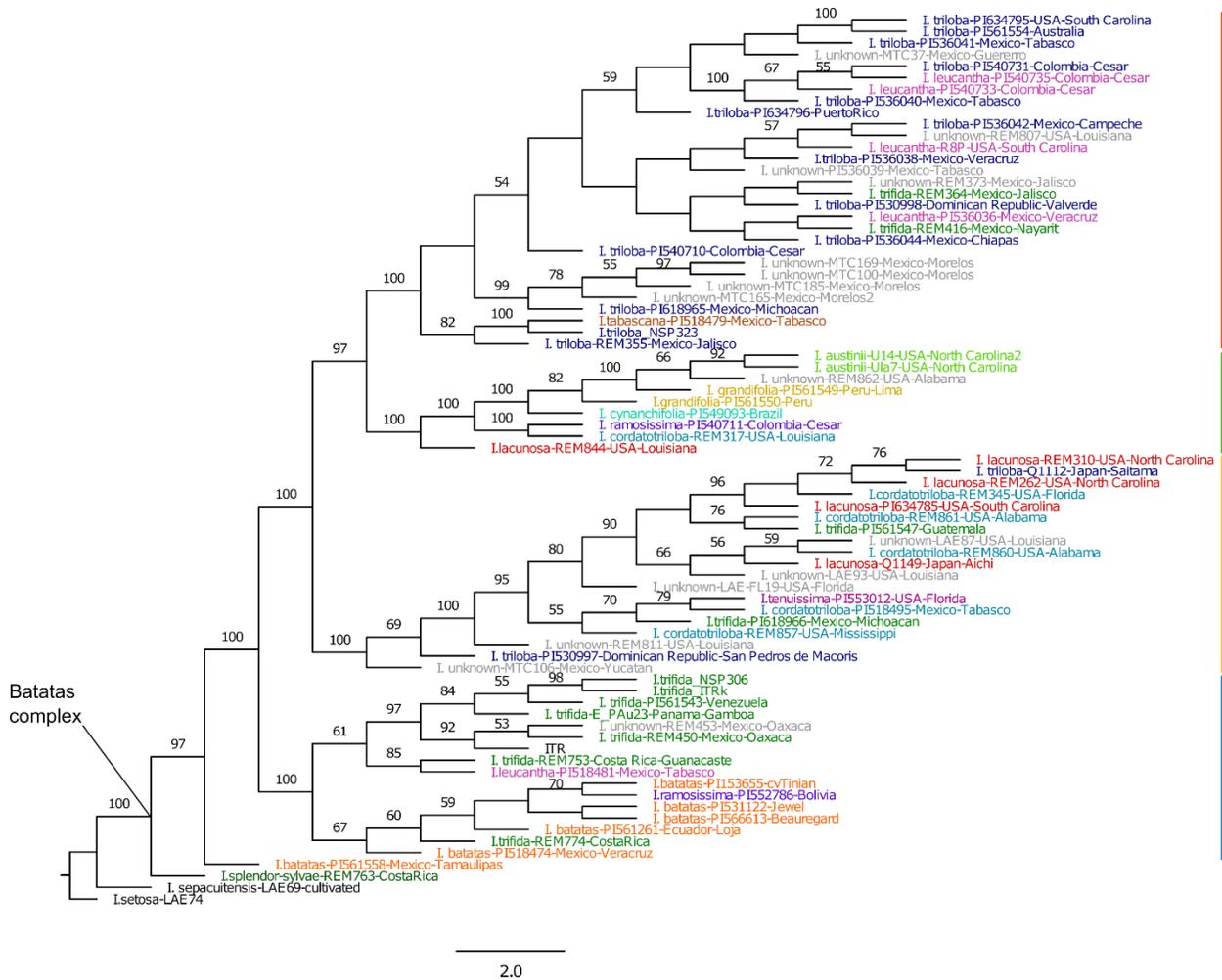


Figure S4.3 – Phylogeny of the Batatas complex estimated in SVDQuartets including diploid and polyploid individuals. Numbers behind nodes indicate bootstrap support from bootstrapping done in SVDQuartets. Nodes without numbers had bootstrap support <50%. Each individual is colored by species designation. Grey individuals are those which exhibited intermediate traits and could not confidently be identified to species using existing taxonomic keys. Colored bars to the right of the phylogeny indicate well-supported clades recovered in the concatenation, ASTRAL-II, and SVDQuartets analyses.

